CANADIAN NETWORK FOR OBSERVATIONAL
DRUG EFFECT STUDIES (CNODES)

Cohort restriction methods to address confounding

(and some propensity score stuff too)

Dan Chateau PhD

CAPT, Toronto

October 24, 2017

CNODES
CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

Disclosure

- none

Other people who I should thank

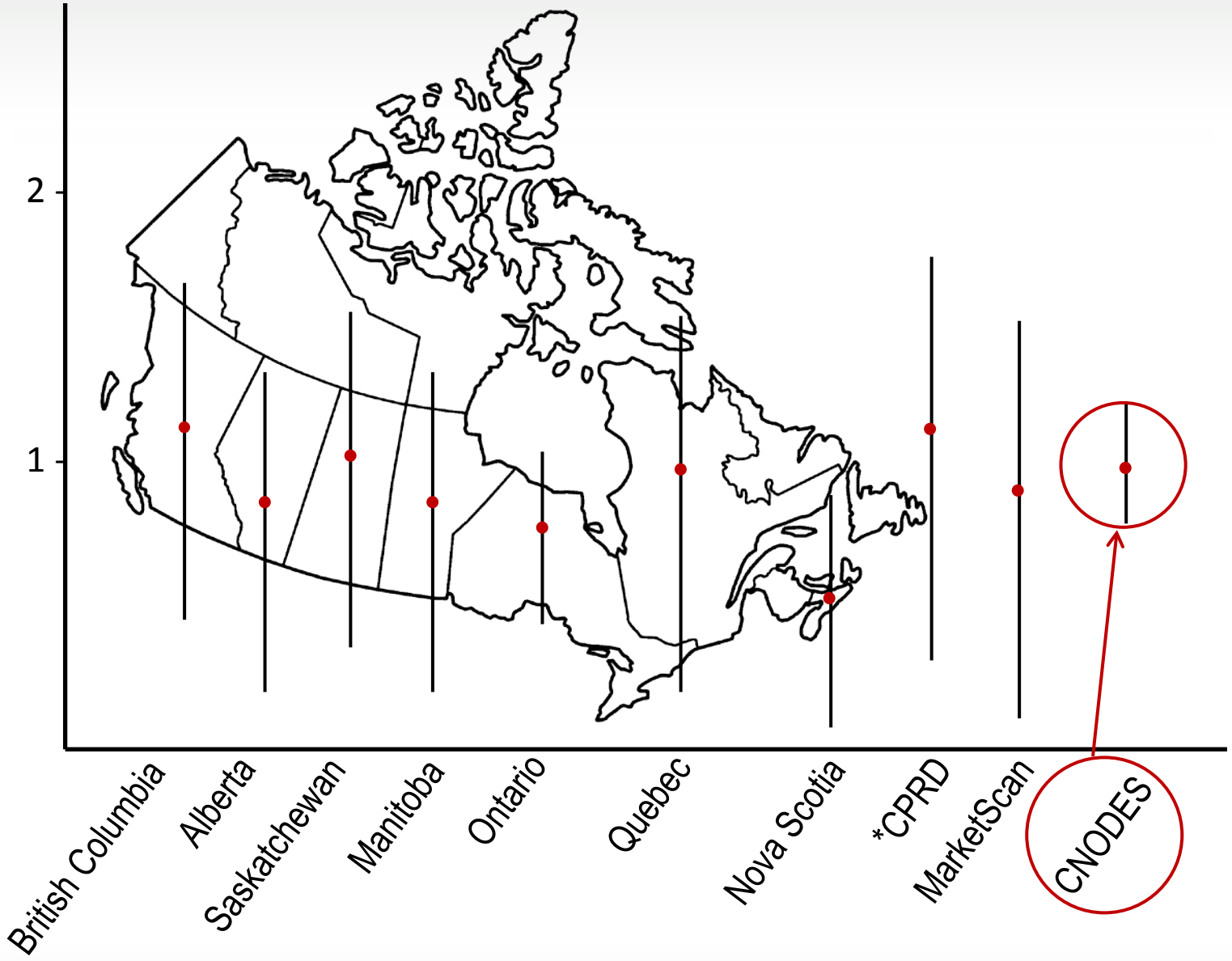
Dan Chateau
Kristian Filion
Colin Dormuth
Robert Platt
Laura Targownik
Matt Dahl
Nathan Nickel

CNODES funding and investigators

Canadian Network for Observational Drug Effect Studies (CNODES), a collaborating centre of the Drug Safety and Effectiveness Network (DSEN), is funded by the Canadian Institutes of Health Research (CIHR, Grant #DSE – 146021).

CNODES INVESTIGATORS	
Executive	Samy Suissa (NPI*), Robert Platt
British Columbia	Colin Dormuth
Alberta	Brenda Hemmelgarn
Saskatchewan	Gary Teare
Manitoba	Patricia Caetano, Dan Chateau
Ontario	David Henry, Michael Paterson
Québec	Jacques LeLorier
Atlantic (NB, NL, NS, PEI)	Adrian Levy, Ingrid Sketris
UK CPRD	Pierre Ernst, Kristian Filion

*Nominated Principal Investigator


2
1
British Columbia
Alberta
Saskatchewan
Manitoba
Ontario
Quebec
Nova Scotia
*CPRD
MarketScan
CNODES

Presentation outline

1. Confounding by indication, protopathic bias

2. Cohort restriction
   - The PPI/HCAP example

3. Some thoughts on propensity scores

Background

- Confounding by indication
  - Indication for drug use is responsible for the drug use and the outcome. Any relationship between the two is spurious.

Background

- Protopathic Bias
  - occurs when a drug treatment is initiated to treat the first symptoms of the disease which is not yet diagnosed

Proton Pump Inhibitors (PPIs) and Hospitalization for Community Acquired Pneumonia (HCAP)

- PPIs
  - main action is a pronounced and long-lasting reduction of gastric acid production
  - one of the most widely sold drugs in the world
  - Nexium, Prevacid, Prilosec

Proton Pump Inhibitors (PPIs) and Hospitalization for Community Acquired Pneumonia (HCAP)

Canadian Medical Association Journal

Eom et al., 2011

- Systematic review and meta-analysis of acid suppressive drugs and risk of pneumonia
- Looked at observational studies (n=8) and RCTs (n=23)

Proton Pump Inhibitors (PPIs) and Hospitalization for Community Acquired Pneumonia (HCAP)

Canadian Medical Association Journal

Eom et al., 2011

- observational studies for PPIs and pneumonia
  - OR = 1.27 (95% CI 1.11 – 1.46) *all*
  - OR = 1.34 (95% CI 1.14 – 1.57) *community acquired only*

Proton Pump Inhibitors (PPIs) and Hospitalization for Community Acquired Pneumonia (HCAP)

Canadian Medical Association Journal

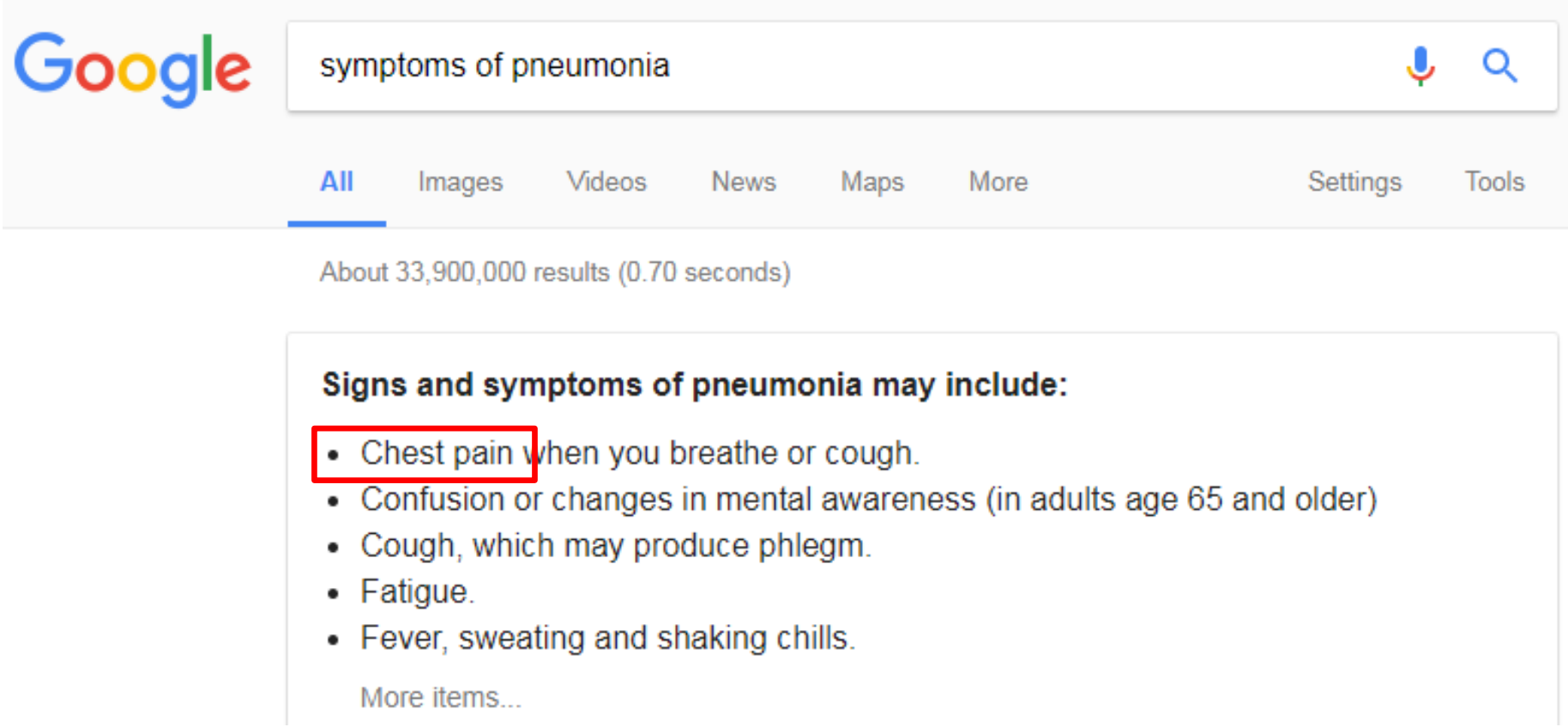
Eom et al., 2011

- Interpretation
  - "Use of a proton pump inhibitor or histamine receptor antagonist may be associated with an increased risk of both community- and hospital –acquired pneumonia

Proton Pump Inhibitors (PPIs) and Hospitalization for Community Acquired Pneumonia (HCAP)

Problems?

- PPIs are prescribed for Gastroesophageal Reflux Disease (GERD), which is an independent risk factor for pneumonia
- protopathic bias - undiagnosed pneumonia leading to prescription of a PPI


Google
symptoms of pneumonia
All
Images
Videos
News
Maps
More
Settings
Tools
About 33,900,000 results (0.70 seconds)
Signs and symptoms of pneumonia may include:
Chest pain when you breathe or cough.
Confusion or changes in mental awareness (in adults age 65 and older)
Cough, which may produce phlegm.
Fatigue.
Fever, sweating and shaking chills.
More items...

symptoms of gerd

All Images News Videos Shopping More Settings Tools

About 24,700,000 results (0.67 seconds)

GERD signs and symptoms include:

- A burning sensation in your chest (heartburn), sometimes spreading to your throat, along with a sour taste in your mouth.
- Chest pain.
- Difficulty swallowing (dysphagia)
- Dry cough.
- Hoarseness or sore throat.
- Regurgitation of food or sour liquid (acid reflux)
- Sensation of a lump in your throat.

Proton Pump Inhibitors (PPIs) and Hospitalization for Community Acquired Pneumonia (HCAP)

Canadian Medical Association Journal

Eom et al., 2011

PPI Duration

<7 days	OR = 3.95 (2.86 – 5.45)
<30 days	OR = 1.61 (1.46 – 1.78)
30-180 days	OR = 1.36 (1.05 – 1.78)

Challenge:

How to construct a cohort study using administrative data that can minimize bias

The team gastroenterologist noted that a small portion of new NSAID prescriptions are combined with a new prophylactic PPI or H2RA prescription

These individuals are unlikely to have undiagnosed pneumonia, and significantly less likely to be suffering from GERD

Results:

There are a lot of new NSAID users (at least 365 days with no NSAID) in the combined data

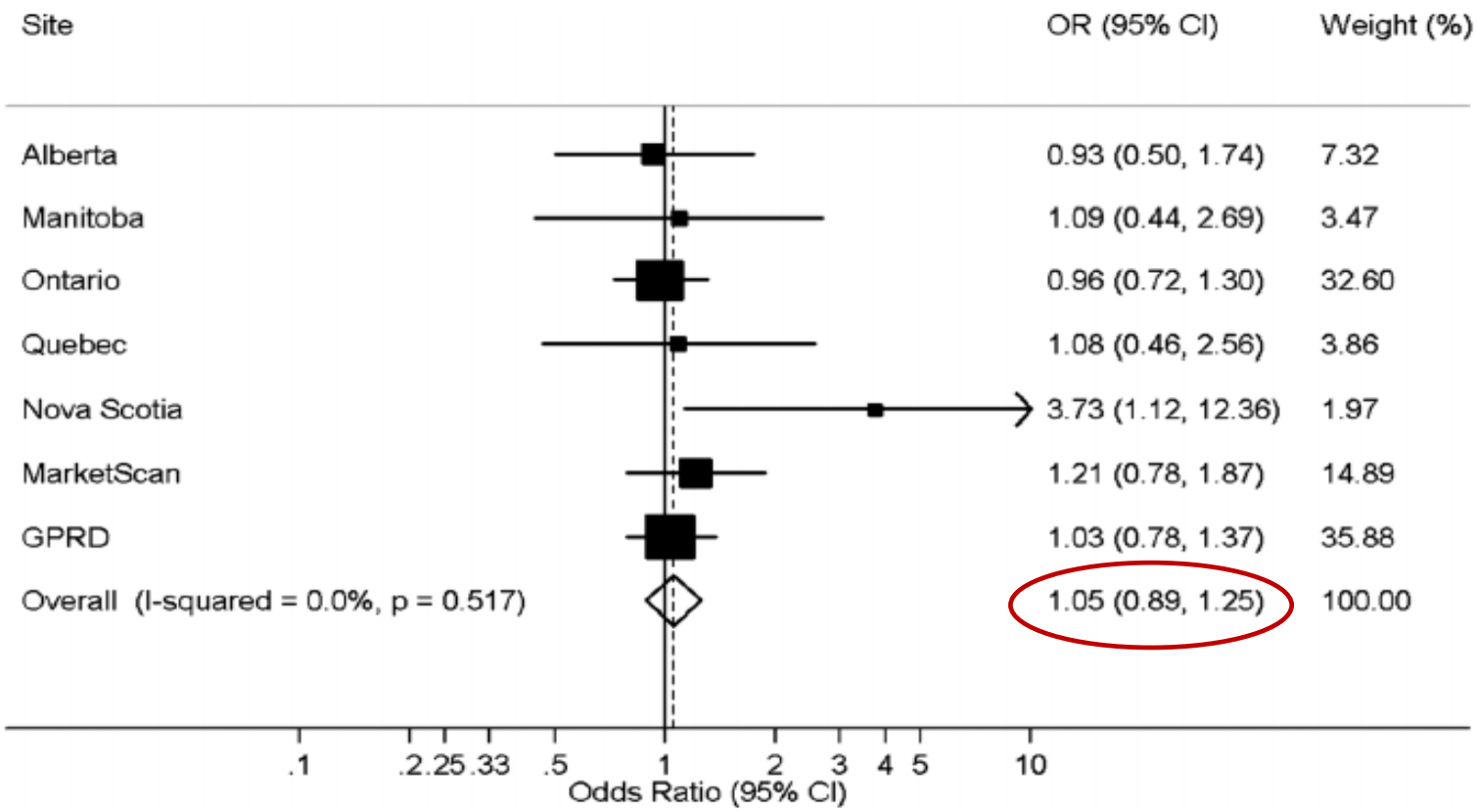
N = ~4.24 million

About 2.3% are concurrently filling a new PPI prescription (at least 365 days with no PPI)

N = ~97000

- *Filion, Chateau et al., Gut. 2014*

  6 month cumulative incidence of HCAP



CNODES
CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

What if we did things the typical way:

MANITOBA results:

	Exposed (PPI User)		Unexposed		Age- and Sex-Adjusted		Fully Adjusted	
Comparison		**No**		**No**	**Odds**	**95% CI**	**Odds**	**95% CI**
	Event	**Event**	**Event**	**Event**	**Ratio**		**Ratio**	
NSAID	192	72506	266	131831	1.27	1.05 - 1.52	1.24	0.96 - 1.59
Statin	297	94377	185	121396	2.01	1.66 - 2.43	1.71	1.29 - 2.26
Antidepressant	304	111455	310	175725	1.15	0.98 - 1.35	1.18	0.96 - 1.46

What if we did things the typical way:

MANITOBA results:

	Exposed (PPI User)		Unexposed (NSAID User)		Age- and Sex-Adjusted		Fully Adjusted	
Exposure Duration	Event	No Event	Event	No Event	Odds Ratio	95% CI	Odds Ratio	95% CI
14 day	29	72669	23	132074	2.21	1.28 - 3.81	2.66	1.29 - 5.48
30 day	41	72657	53	132044	1.35	0.90 - 2.03	1.64	0.97 - 2.79
90 day	104	72594	142	131955	1.28	0.99 - 1.65	1.19	0.85 - 1.68

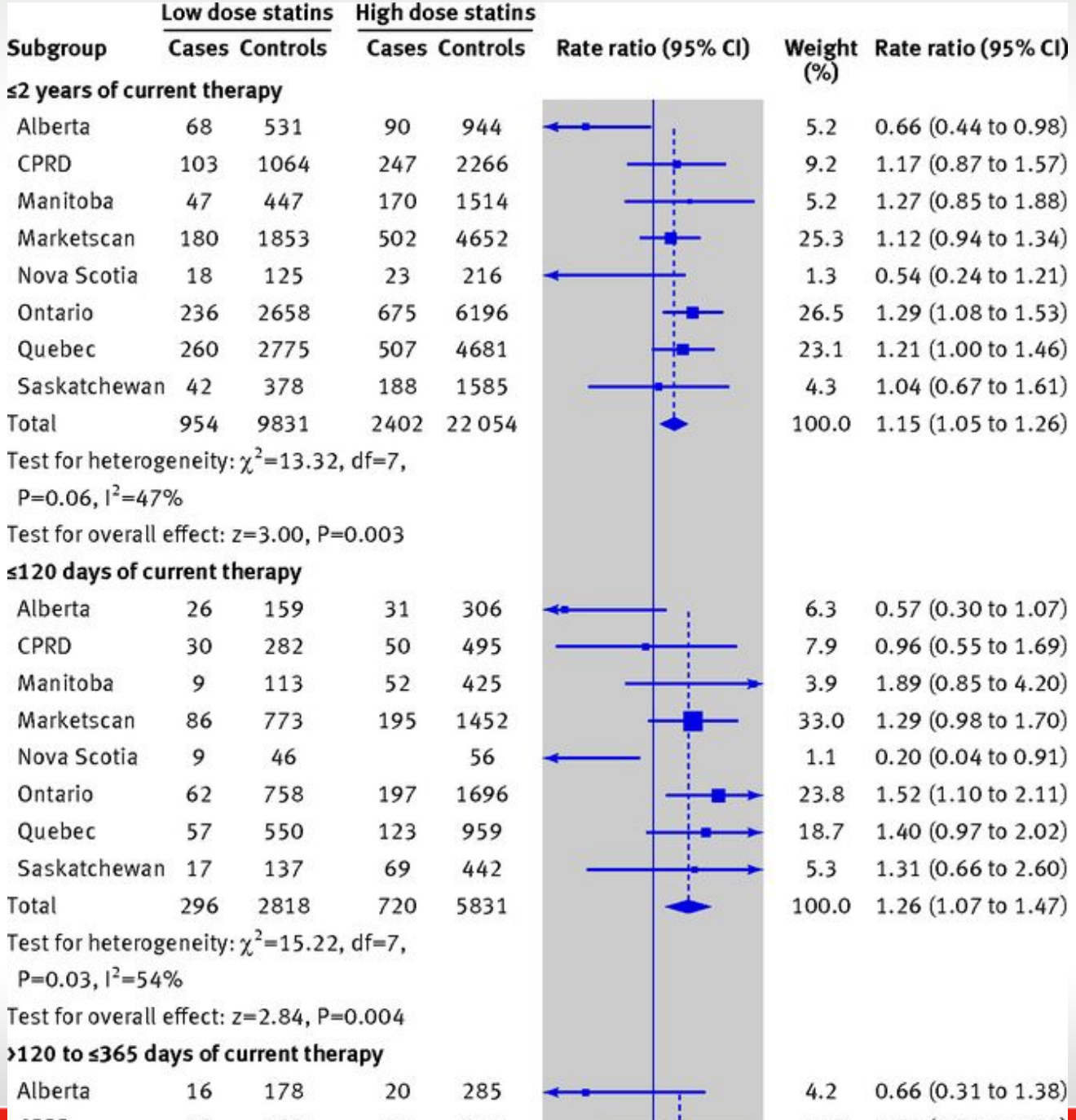
Statins and Diabetes

- FDA Advisory
  - Confounding by indication
  - Prevalent diabetes needs to be removed
- Study patients admitted to hospital for a CV event where diabetes was not reported in the DAD (discharge abstract)

Statins and Diabetes

- Secondary prevention cohort
  - Received a new statin within 90 days of hospital discharge (new user with 1 year washout)
  - Diabetes defined as occurrence of a subsequent hospital discharge with a diabetes diagnosis or a new prescription for insulin or oral antidiabetic medication

Colin R Dormuth et al. BMJ 2014;348:bmj.g3244



Subgroup	Low dose statins Cases	Low dose statins Controls	High dose statins Cases	High dose statins Controls	Rate ratio (95% CI)	Weight (%)	Rate ratio (95% CI)
≤2 years of current therapy							
Alberta	68	531	90	944		5.2	0.66 (0.44 to 0.98)
CPRD	103	1064	247	2266		9.2	1.17 (0.87 to 1.57)
Manitoba	47	447	170	1514		5.2	1.27 (0.85 to 1.88)
Marketscan	180	1853	502	4652		25.3	1.12 (0.94 to 1.34)
Nova Scotia	18	125	23	216		1.3	0.54 (0.24 to 1.21)
Ontario	236	2658	675	6196		26.5	1.29 (1.08 to 1.53)
Quebec	260	2775	507	4681		23.1	1.21 (1.00 to 1.46)
Saskatchewan	42	378	188	1585		4.3	1.04 (0.67 to 1.61)
Total	954	9831	2402	22 054		100.0	1.15 (1.05 to 1.26)
Test for heterogeneity: χ^2=13.32, df=7, P=0.06, I^2=47%							
Test for overall effect: z=3.00, P=0.003							
≤120 days of current therapy							
Alberta	26	159	31	306		6.3	0.57 (0.30 to 1.07)
CPRD	30	282	50	495		7.9	0.96 (0.55 to 1.69)
Manitoba	9	113	52	425		3.9	1.89 (0.85 to 4.20)
Marketscan	86	773	195	1452		33.0	1.29 (0.98 to 1.70)
Nova Scotia	9	46		56		1.1	0.20 (0.04 to 0.91)
Ontario	62	758	197	1696		23.8	1.52 (1.10 to 2.11)
Quebec	57	550	123	959		18.7	1.40 (0.97 to 2.02)
Saskatchewan	17	137	69	442		5.3	1.31 (0.66 to 2.60)
Total	296	2818	720	5831		100.0	1.26 (1.07 to 1.47)
Test for heterogeneity: χ^2=15.22, df=7, P=0.03, I^2=54%							
Test for overall effect: z=2.84, P=0.004							
>120 to ≤365 days of current therapy							
Alberta	16	178	20	285		4.2	0.66 (0.31 to 1.38)

Another approach not tested…

- Monomaniac prescribers

- Study only patients treated by this subset of physicians

Propensity Scores

- High Dimensional Propensity Score algorithm
- Inverse Probability of Treatment Weights or IPTWs

FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

- Cannot Observe…

  The same person under both conditions

Average Effects

- Effects vary from individual to individual

- Average effect tells us…

  “the effect for a person--at random-- from our group.”

How do we get this “average effect”?

Compare average outcomes

Average Effects

- One state: the group **gets** the drug
- One state: the group **does not receive** the drug
- Difference in average outcomes between the two is the average effect of the drug

FUNDAMENTAL PROBLEM OF CAUSAL INFERENCE

CANNOT OBSERVE GROUP under BOTH STATES

- Groups that we can actually observe
  - Those that really did receive the drug
  - Those that did not receive the drug
- Ask: Are these two groups comparable?
  - Minimize possibility that observed differences are due to confounding
- Strategies to deal with confounding:
  - Multiple Regression
  - Matching
  - **Propensity Score Methods**

The Propensity Score—Review

- The Propensity Score: Probability that person is exposed:
  The probability that the person receives the drug

The Propensity Score--Review

- If probability (aka propensity score) is close to 1
  - —VERY LIKELY to receive the drug given observed covariates
- If probability (aka propensity score) is close to 0
  - —VERY UNLIKELY to receive the drug given observed covariates

- Use propensity score to make and compare comparable groups

Epidemiology. 2009 Jul;20(4):512-22. doi: 10.1097/EDE.0b013e3181a663cc.

High-dimensional propensity score adjustment in studies of treatment effects using health care claims data.

Schneeweiss S[1], Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA.

⊕ Author information

Abstract

BACKGROUND: Adjusting for large numbers of covariates ascertained from patients' health care claims data may improve control of confounding, as these variables may collectively be proxies for unobserved factors. Here, we develop and test an algorithm that empirically identifies candidate covariates, prioritizes covariates, and integrates them into a propensity-score-based confounder adjustment model.

METHODS: We developed a multistep algorithm to implement high-dimensional proxy adjustment in claims data. Steps include (1) identifying data dimensions, eg, diagnoses, procedures, and medications; (2) empirically identifying candidate covariates; (3) assessing recurrence of codes; (4) prioritizing covariates; (5) selecting covariates for adjustment; (6) estimating the exposure propensity score; and (7) estimating an outcome model. This algorithm was tested in Medicare claims data, including a study on the effect of Cox-2 inhibitors on reduced gastric toxicity compared with nonselective nonsteroidal anti-inflammatory drugs (NSAIDs).

RESULTS: In a population of 49,653 new users of Cox-2 inhibitors or nonselective NSAIDs, a crude relative risk (RR) for upper GI toxicity (RR = 1.09 [95% confidence interval = 0.91-1.30]) was initially observed. Adjusting for 15 predefined covariates resulted in a possible gastroprotective effect (0.94 [0.78-1.12]). A gastroprotective effect became stronger when adjusting for an additional 500 algorithm-derived covariates (0.88 [0.73-1.06]). Results of a study on the effect of statin on reduced mortality were similar. Using the algorithm adjustment confirmed a null finding between influenza vaccination and hip fracture (1.02 [0.85-1.21]).

CONCLUSIONS: In typical pharmacoepidemiologic studies, the proposed high-dimensional propensity score resulted in improved effect estimates compared with adjustment limited to predefined covariates, when benchmarked against results expected from randomized trials.

Eur J Clin Pharmacol. 2016 Dec;72(12):1497-1505. Epub 2016 Aug 30.

Performance of the high-dimensional propensity score in adjusting for unmeasured confounders.

Guertin JR[1,2,3], Rahme E[4,5], LeLorier J[6].

Author information

Abstract

PURPOSE: High-dimensional propensity scores (hdPS) can adjust for measured confounders, but it remains unclear how well it can adjust for unmeasured confounders. Our goal was to identify if the hdPS method could adjust for confounders which were hidden to the hdPS algorithm.

METHOD: The hdPS algorithm was used to estimate two hdPS; the first version (hdPS-1) was estimated using data provided by 6 data dimensions and the second version (hdPS-2) was estimated using data provided from only two of the 6 data dimensions. Two matched sub-cohorts were created by matching one patient initiated on a high-dose statin to one patient initiated on a low-dose statin based on either hdPS-1 (Matched hdPS Full Info Sub-Cohort) or hdPS-2 (Matched hdPS Hidden Info Sub-Cohort). Performances of both hdPS were compared by means of the absolute standardized differences (ASDD) regarding 18 characteristics (data on seven of the 18 characteristics were hidden to the hdPS algorithm when estimating the hdPS-2).

RESULTS: Eight out of the 18 characteristics were shown to be unbalanced within the unmatched cohort. Matching on either hdPS achieved adequate balance (i.e., ASDD <0.1) on all 18 characteristics.

CONCLUSION: Our results indicate that the hdPS method was able to adjust for hidden confounders supporting the claim that the hdPS method can adjust for at least some unmeasured confounders.

Weight Population to Estimate Treatment Effects

- Analytic Sample: Everyone Eligible to receive the drug:

- Treatment Effects

  - What is the average effect of the drug among all of the people who could get it?

  - What is the effect of the drug among the people who ACTUALLY received it?

Weights: ATE

- **Average Treatment Effect** :

  Imagine we take EVERYONE in the TARGET POPULATION…

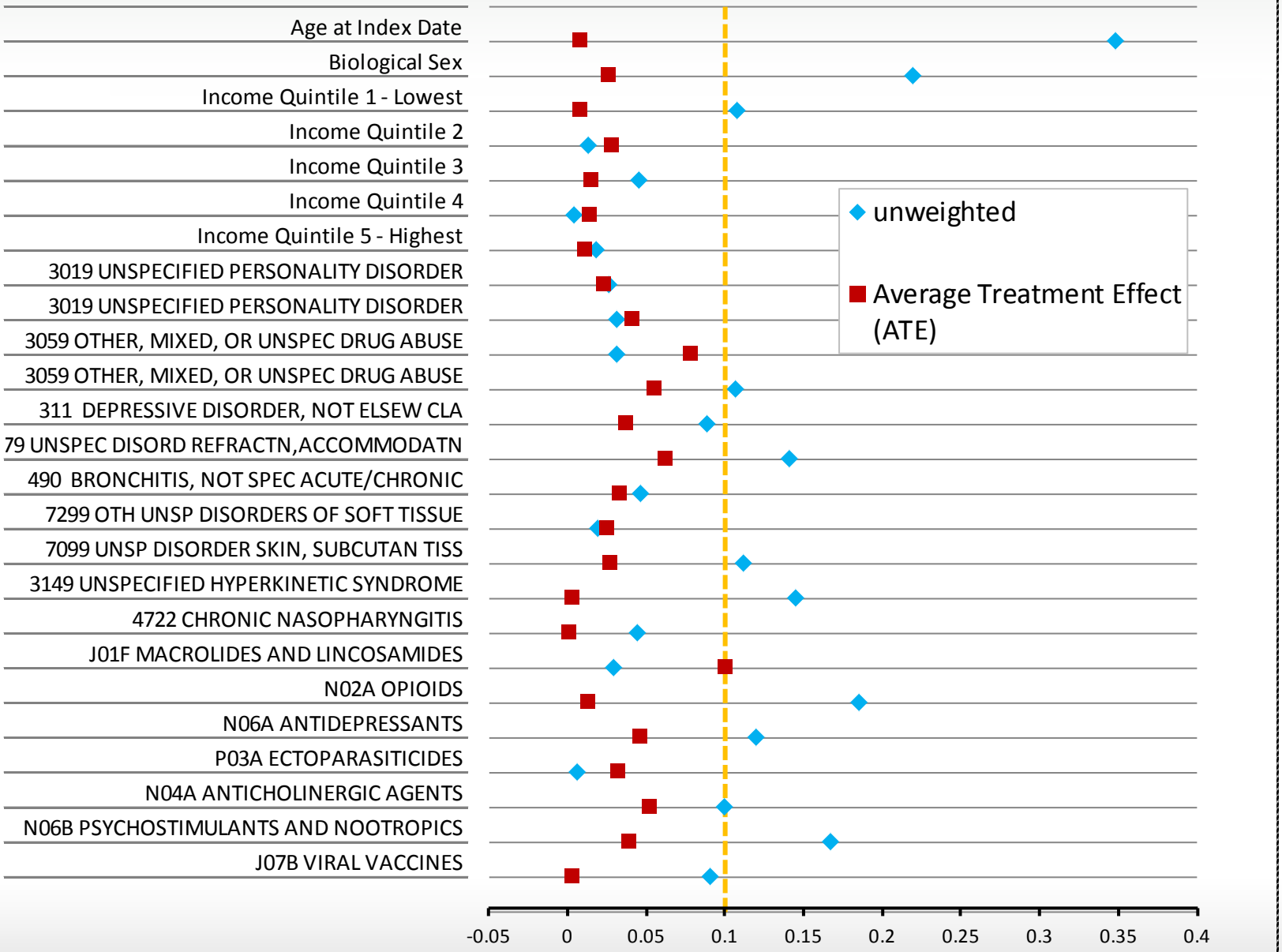
  - EVERYONE **receives** the drug

    COMPARED WITH…

  - EVERYONE **does not receive** the drug

$$WEIGHT_{Average\ Treatment\ Effect} = FDK * \frac{1}{PS} + (1 - FDK) * \frac{1}{1 - PS}$$

Standardized differences



Thank you

Visit us at www.cnodes.ca

CNODES

CANADIAN NETWORK FOR OBSERVATIONAL DRUG EFFECT STUDIES

CIHR IRSC

New Method for Imputing Missing Confounders, based on a Small Validation Sample, in Time-to-Event Analyses

Michal Abrahamowicz

James McGill Professor
Department of Epidemiology & Biostatistics
McGill University

& Rebecca Burne (McGill)

Support: CAN-AIM grant from the Drug Safety or Effectiveness Network (DSEN)
of the *Canadian Institutes for Health Research (CIHR)*

McGill

CAN-AIM

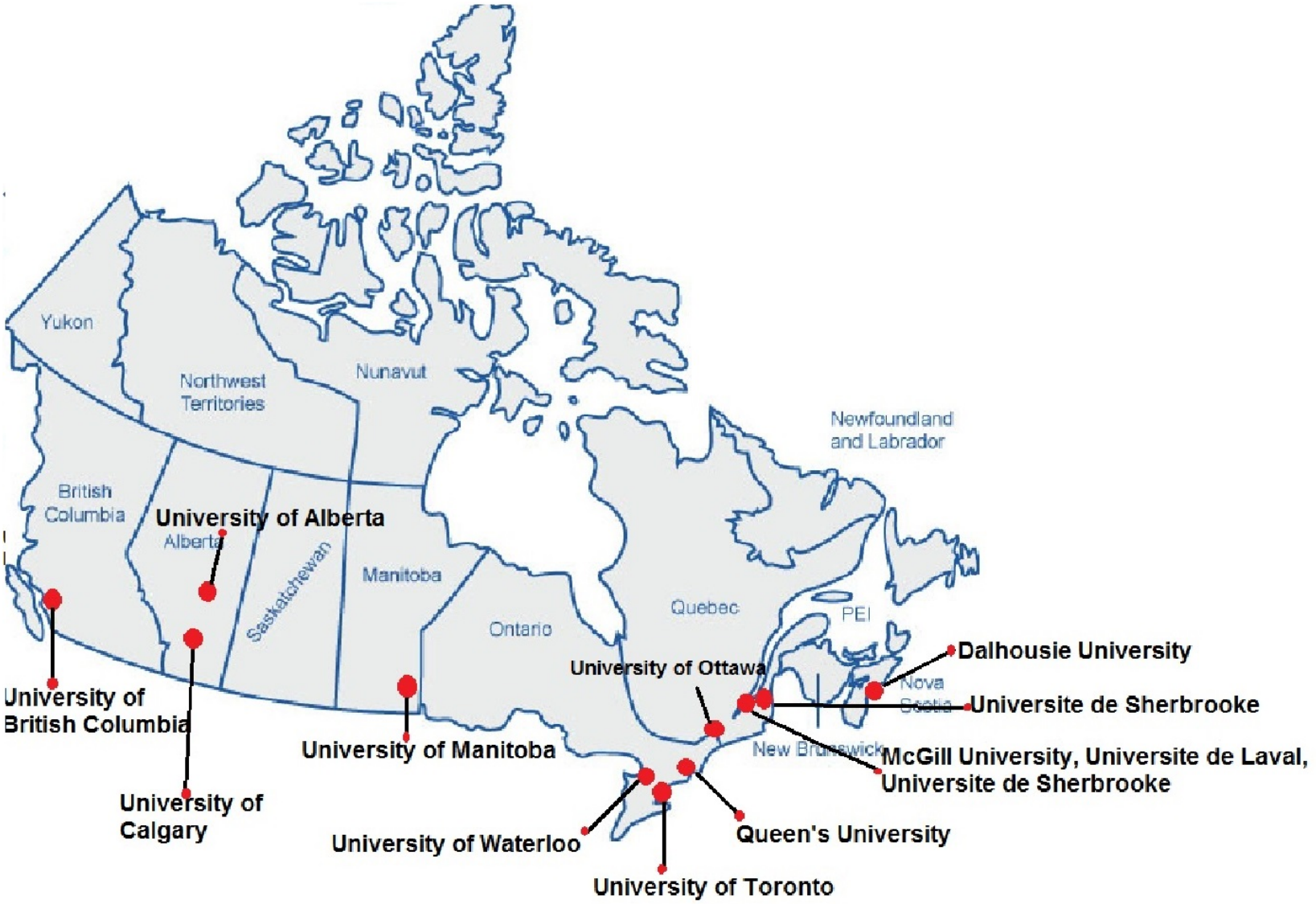
Objectives of the CAN-AIM project

- CAnadian Network for Advanced Interdisciplinary Methods for comparative effectiveness research (CAN-AIM) network is a pan-Canadian network of >35 researchers from 12 universities, in 6 provinces, funded by **the Drug Safety or Effectiveness Network (DSEN) of the CIHR** to:

  1) **Develop and Validate new Methods for improving analyses of population-based studies of Safety and Effectiveness of Drugs, with focus on Prospective Studies (time-to-event analyses)**

  2) Apply these new methods to address Queries from Health Canada (and provincial agencies) regarding safety or effectiveness of specific drugs used by Canadians

CAN-AIM map


Yukon
Northwest Territories
Nunavut
Newfoundland and Labrador
British Columbia
Alberta
Saskatchewan
Manitoba
Ontario
Quebec
PEI
Nova Scotia
New Brunswick
University of Alberta
University of British Columbia
University of Calgary
University of Manitoba
University of Ottawa
Dalhousie University
Universite de Sherbrooke
McGill University, Universite de Laval, Universite de Sherbrooke
Queen's University
University of Waterloo
University of Toronto

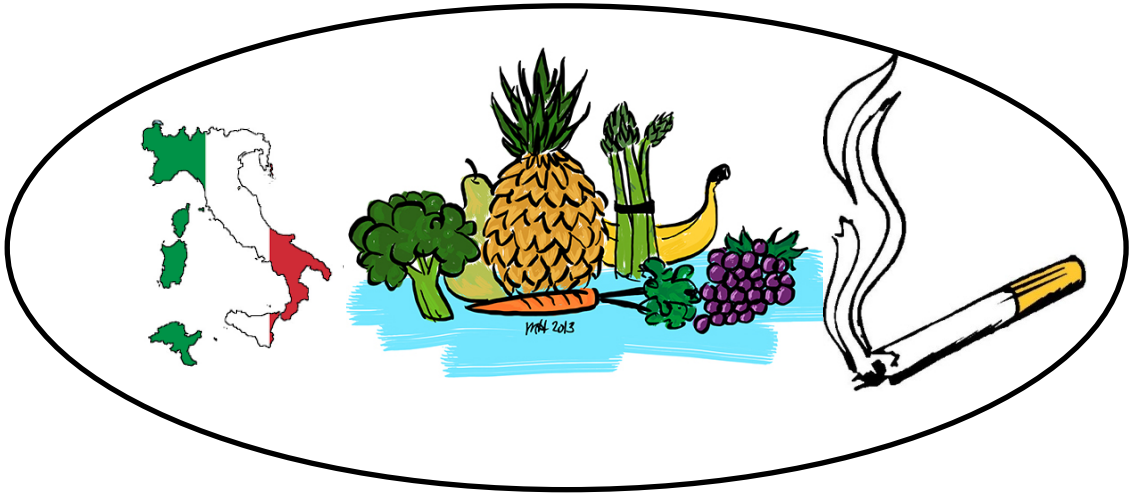
Confounding: Basic Concepts

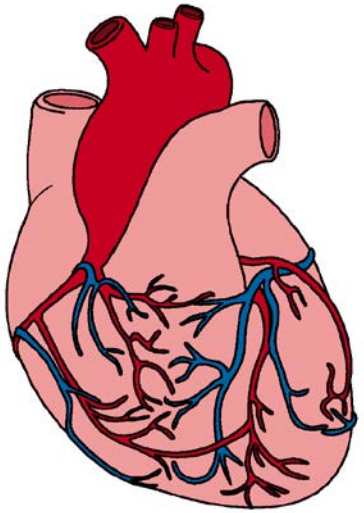
- **Confounding of the estimated Association between 'Exposure' (X) and 'Outcome' (Y) occurs due to a failure to account for Another Variable ('Confounder'), Associated (causally or 'just by chance') with Both Exposure & Outcome**

- **Un-Measured (un-observed) Confounding** is a major problem in Observational, Population-based studies of Drug Safety or Effectiveness where some potential **Confounder(s) 'U' are Not Recorded** i.e. un-measured (common in Large Administrative Database studies)
  [Walker *Epidemiology* 1996; Wolfe et al *J Rheum* 2002]

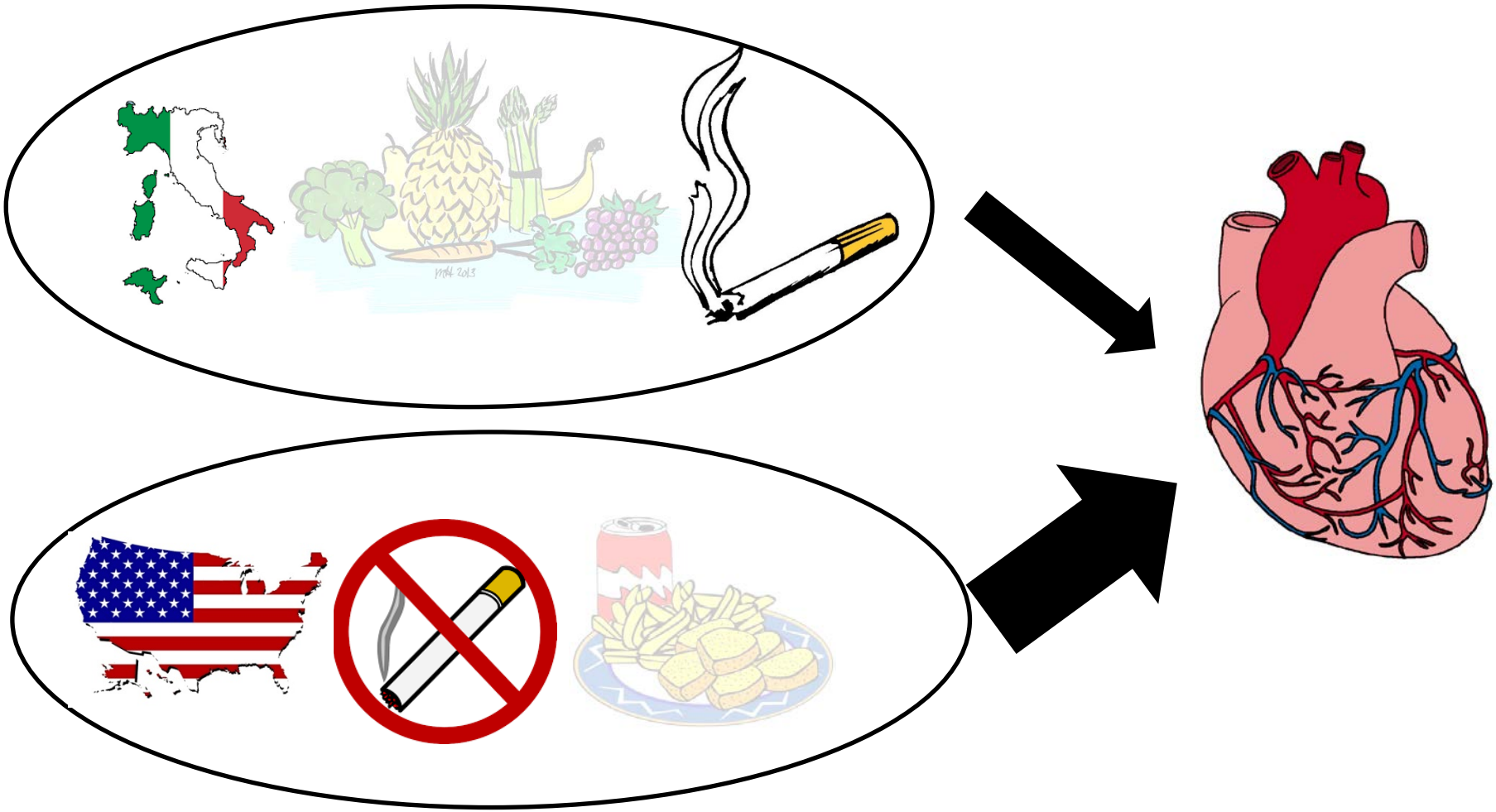
- Then, the **Estimated X ⇒ Y association will be Biased** with the **Direction and Severity of the Bias** depending on the pattern of U ⇔ X and U ⇒ Y associations**
  (** e.g.,
  (i) inducing a Spurious association = Type I Error;
  (ii) Under- or Over-estimation of its strength; or
  (iii) in extreme cases: Reversal of the direction of the Association)

Southern Europe (healthy Mediterrean diet + ~40% smoking-
⇒ ~ 200 CVD deaths/100 K)
versus North America (fast food + ~ 20% smoking
⇒ ~ 400 CVD deaths/100 K))





If DIET is NOT Measured (or accounted for),
Smoking will appear ‘Protective’ against Coronary Disease



Confounding by Indication

- **Confounding by Indication is a Major Source of Bias in Observational Pharmaco-Epidemiological Studies of Drugs Effectiveness/Safety**

  [Walker et al, *Epidemiology* 1996; Avorn, *NEJM* 2007; Patorno et al, *PDS* 2015]

- Such Studies usually rely on **Large Administrative Databases** (to ensure adequate Power & Precision) [Skegg et al, *Stat Med* 2001], that d**o NOT provide information on such Important Risk Factors** as e.g. Disease Severity, Smoking, Obesity, Blood Pressure or Lab Tests/Biomarkers

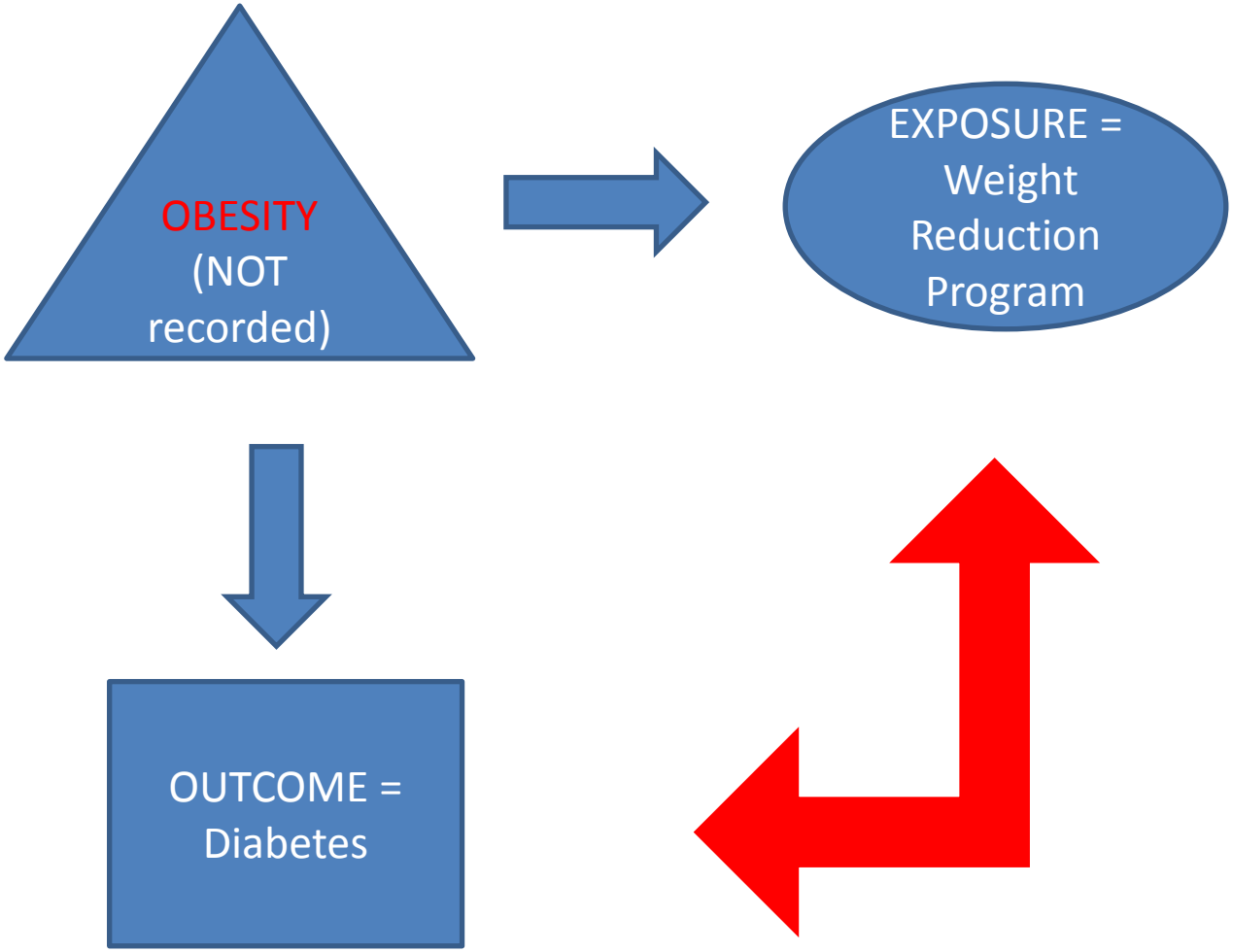
  [e.g., Wolfe et al, *J Rheum* 2002]

- Yet, in real-life Clinical Practice, (**i) Choice of Drugs or Treatments depends on several Clinical and Socio-demographic patient's characteristics, that (ii) may also Affect the Outcome**, i.e. **Act as Potential Un-measured Confounders of the Treatment Effect**

Sources of Confounding by Indication in Pharmaco-epidemiology

- (A) People are prescribed Drugs because they have a Disease or are at Increased Risk of Developing a Disease
- (A) $\Rightarrow$ (B) : those Prescribed a Drug will be generally Sicker/more Vulnerable and, thus, will have Worse Outcomes than those who are Un-treated
- (C) Large Admin Databases do NOT record Disease Severity and many Clinical Risk Factors
- {(B) & (C)} $\Rightarrow$ **spurious 'evidence' of an Association between Use of Drugs and Bad Outcomes (High Risk)**
- **However, this "Association" in fact reflects the 'Reversal Causality Bias': patients receive drugs Because of being at high risk (NOT *vice versa*)**
- **Unmeasured Confounding by Indication = Harmful Effects of Drugs/Treatments/Interventions**

"be careful about ***Adverse Effects of these Weight Watching programs****: we noticed that* ***many patients diagnosed with Diabetes have recently started one…"***



Need to Develop & Validate Alternative Methods
to deal with Unmeasured Confounding

- **Unmeasured Confounding is a Complex Phenomenon** and both its **Sources & Impact vary considerably depending on the study design, data structure, data availability etc.**
- Very **Unlikely to Ever find a *Panaceum* solution** ("*one method fits all*")
- Thus: **Alternative Methods have to be Developed** for Different (relevant) Situations and **under Different (Plausible) Assumptions**
- **It is Essential to Validate New Methods** & Systematically Assess and Compare their Performance, ideally **through Simulation Studies** (where the Estimated Effects are validated against the Known 'True' Effects)

Overview of Existing Analytical Methods for Unmeasured Confounding in **Pharmaco-epi**

- General Lack of Appropriate Methods* until early 2000's
  [review by McMahon, *Pharmacoepi & Drug Safety* 2003]

- * Exception: Bias Sensitivity Analyses:
  (i) assume **Hypothetical Confounder(s)** U with specific U $\Leftrightarrow$ X & U $\Rightarrow$ Y associations, then (ii) assess, through Analysis [Greenland, *Int J Epi* 1996] or Simulations [Groenwold et al, *Int J Epi* 2010] how adjusting for U affects the estimated X $\Rightarrow$ Y association
  (example of Application: [Pilote et al, *Ann Int Med* 2004])

- Recent **High Dimension Propensity Score (hdPS)** method:
  [Schneeweiss et al, *Epidemiology* 2009]
  (example of Application: [Kumamaru et al, *J Clin Epidemiol* 2016]):
  hdPS approach was NOT Systematically Evaluated

Overview of Existing Analytical Methods for Unmeasured Confounding in **Pharmaco-epi (Cont-d)**

- Instrumental Variables (IV) approach based on Physician Prescribing Preferences [Brookhart et al, *Epidemiology* 2006]:
  - Advantage: if underlying assumptions are correct, the IV approach **Removes Bias due to Unobserved Confounding by Indication** [Brookhart et al 2006; Abrahamowicz et al 2011];
  - Limitations:
    - Depends critically on s**everal assumptions** [Brookhart et al 2006];
    - Serious **VARIANCE INFLATION** [Ionescu-Ittu et al 2009; 2012]
    - **Difficult to Adapt to Time-to-Event analyses**
- **The "Missing Cause" approach** [Abrahamowicz et al, *Stat Med* 2016]
  - Assumptions similar to IV approach
  - Not yet evaluated for time-to-event analyses

Conclusions

- Each of the existing methods relies on important assumptions and is applicable only in specific situations
- Only very few methods have been developed and validated for Time-to-Event analyses that are essential for prospective (or retrospective) cohort studies
- Need to Develop & Validate NEW METHODS

Background: Administrative Data

- Pharmacoepidemiologic studies [Skegg, *Stat Med* 2001]:
  - Rely on **Administrative Databases** ("Main database")
  - Data collected routinely/electronically (e.g. for administration of health services / insurance)

- <u>Advantages:</u>
  - **Large N & Long follow-up** (□**Power** to detect even weak associations)
  - **Population-based (No selection bias)**
  - **Complete information on drug prescriptions**, major comorbidities, hospital admissions, health outcomes…

Disadvantages of Administrative Data

- Data *not* collected for research purposes
- **Often no measurements on important potential confounders [1] e.g.:**
  - Lifestyle characteristics
  - Laboratory tests
  - Disease severity [1]
- **Unmeasured confounding is an important problem! [2, 3, 4]**

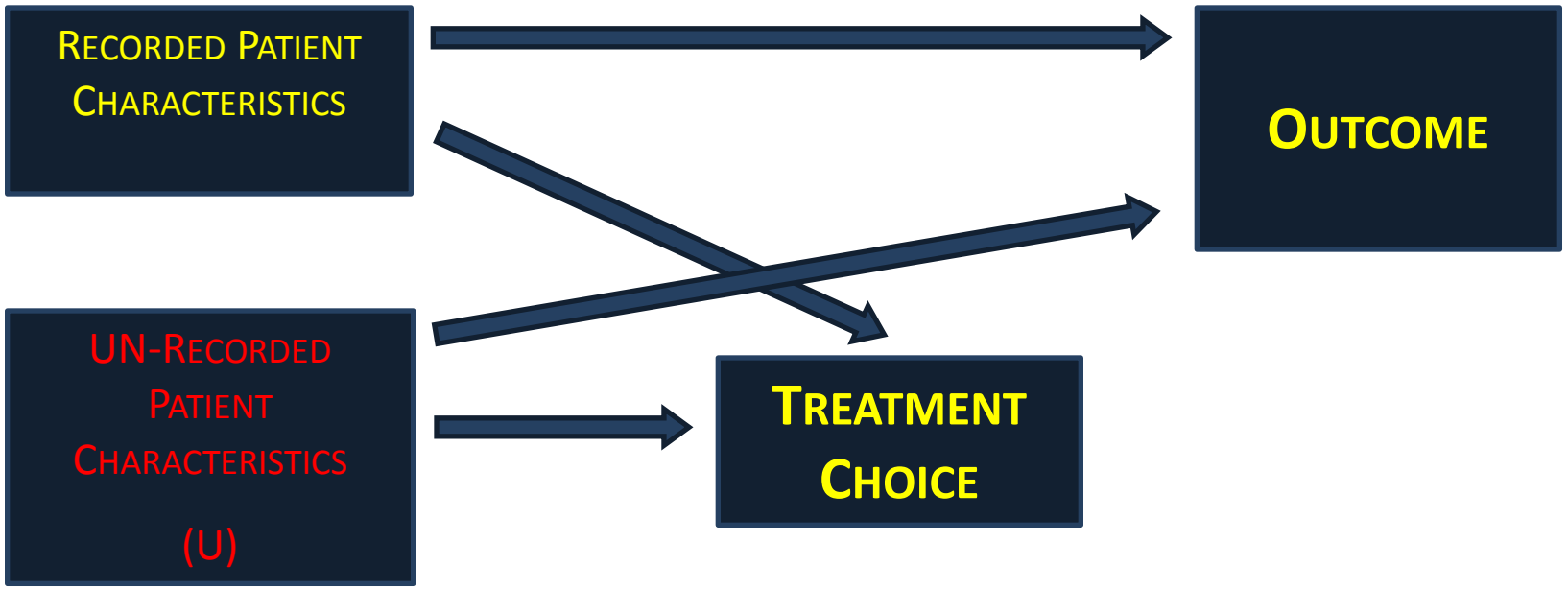
[1] Wolfe et al, *The Journal of Rheumatology* 2002

[2] Patorno et al, *PDS* 2015

[3] Avorn, *NEJM* 2007

[4] Walker, *Epidemiology* 1996

Conceptual Framework: Determinants of Treatment Choices



Clinical datasets – a solution?

- **some important additional confounders may be recorded in a clinical dataset ("Validation subsample"= VS)** [Sturmer et al, *Am J Epidemiology* 2005]
- VS data are collected for research purposes

  (e.g. prospective cohort following persons with a specific disease, & records also drug use & clinical outcomes of interest)

  – Advantage:

  Measurement of important confounders (**Avoids bias**)

  – Disadvantage:

  Smaller size (**Inadequate power & precision**)

Bias/Variance trade-off in Pharmaco-Epi

- **Large Administrative Database studies** ensure adequate Power and Precision, but fail to measure important Confounders, leading to **BIAS** in the Treatment effect estimates
- Research-oriented **Clinical Cohort studies** provide data on Confounders un-measured in the large studies but their Small Sample Size, combined with Rare Events, **Inflate VARIANCE,** resulting in Insufficient Power and Imprecise Estimates

Proposed Solution

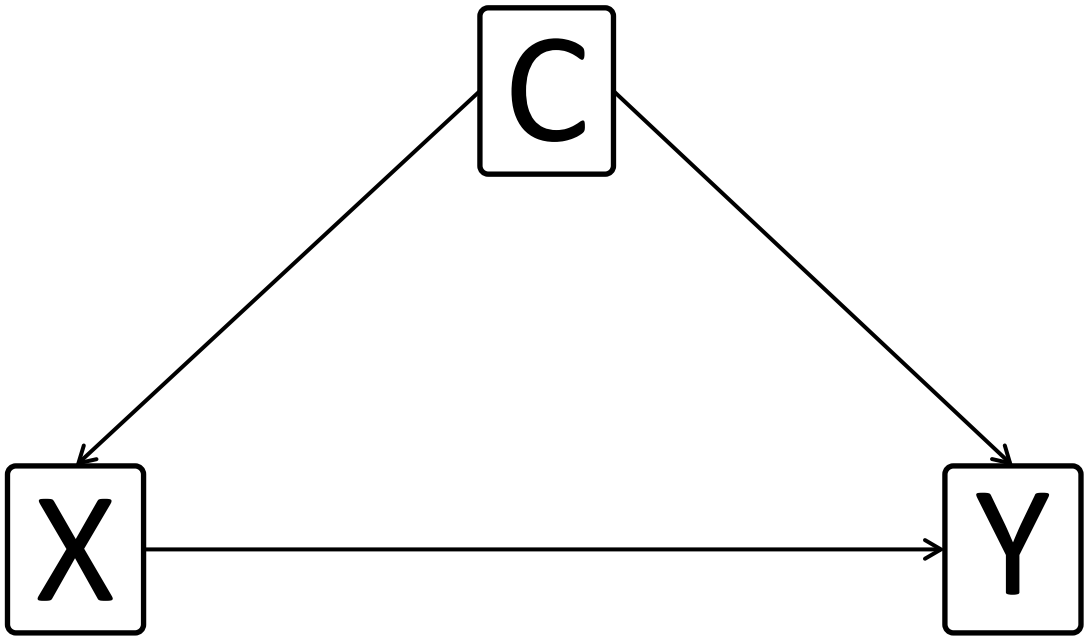
- **We Develop New Statistical Methods in order to COMBINE the STRENGTHS of the 2 Data Sources:**

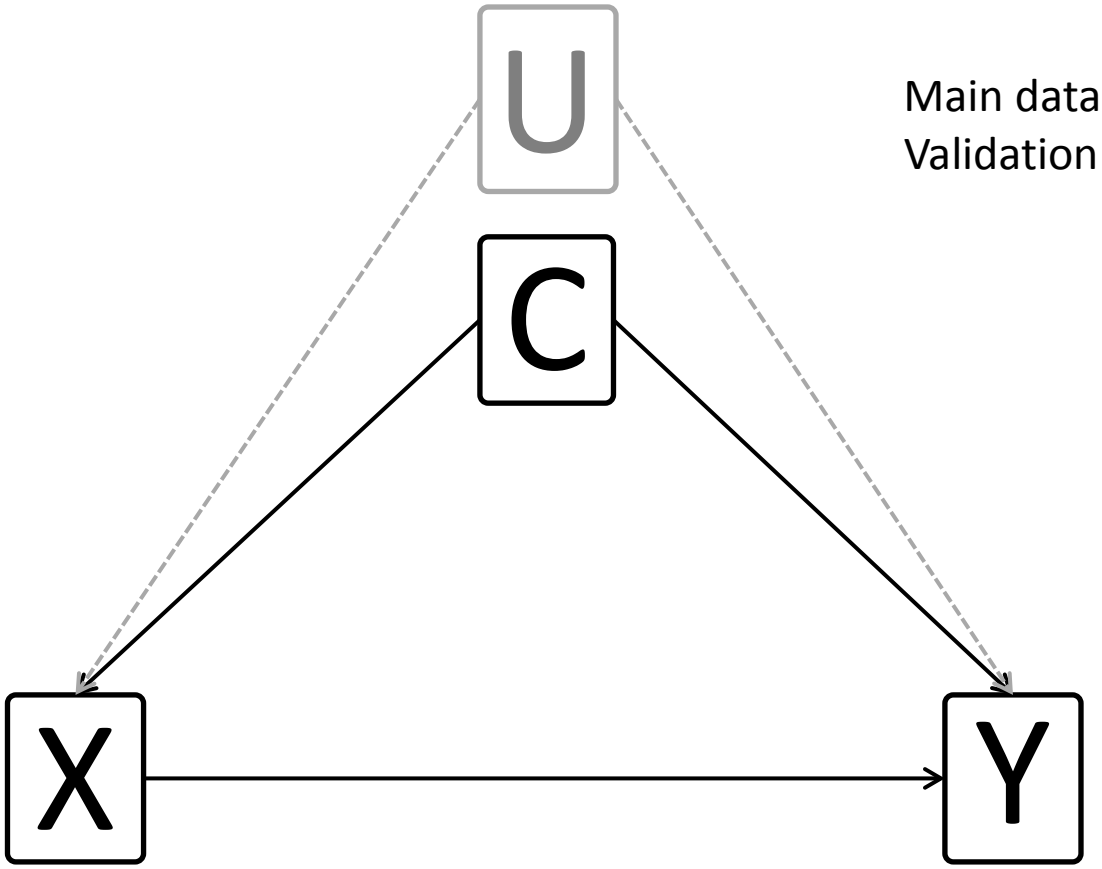
  (1) **High Power/Precision** of the **Large Administrative Databases**

  &

  (2) **Adequate Control for Confounders** measured only in **Clinical Cohort studies (with Small effective Sample Sizes)**, used as **"Validation Samples" (VS)**

Main database: {**C**,X,Y}





Main database: {**C**,X,Y}
Validation subsample: {**C**,**U**,X,Y}

Current methods to use VS data

Current methods, which use additional data on potential confounders available only in the smaller **Validation Sample to account for unmeasured confounding** in the main (large) database include:

1. ***Propensity Score Calibration*** **(PSC)**
   [Stürmer et al, *Am J Epidemiol* 2005]

2. ***Bayes PS***
   [McCandless et al, *JASA* 2012]

- **Only PSC extended to Time-to-Event analyses** but **requires strong "surrogacy assumption"**
  [Stürmer et al, *Am J Epidemiol* 2007]

Propensity Score Calibration (PCS): 4-step Implementation [Sturmer et al, *AJE* 2005]

1) In the Validation Sample (VS) estimate 2 versions of Propensity Score (PS):
   i. 'Error Prone' PS_{EP} (adjusted Only for confounders measured in the main database)
   ii. 'Gold Standard' PS_{GS} (adjusted for All confounders, including those measured in the VS only)
2) In the VS: fit the linear regression of PS_{GS} on PS_{EP}
3) In the Main Database: estimate PS_{GS} for each subject, based on the model fit in step 2 (as a function of PS_{EP})
4) In the Main Database: fit the outcome model using PS_{GS} estimates from step 3 for adjustment or matching

PSC: Surrogacy Assumption

- PSC relies on surrogacy: PS_{EP} independent of outcome conditional on PS_{GS} and exposure
- Holds when direction of confounding of **U** and **C** is the same
- Adding an unmeasured confounder to the PS will increase the strength of the association in same direction
- If directions differ, including an unmeasured confounder in PS will decrease the strength of the association, therefore, PS_{EP} has stronger association with outcome than PS_{GS}

Direction of confounding = direction in which exposure-outcome relationship is biased if the confounder is not adjusted for

Need to Extend the current methods

a) **Time-to-event (survival) analyses:**
   - Both PSC and BayesPS have been validated for **binary outcomes** only, not time-to-event analyses,
   - Yet, pharmacoepi studies based on administrative databases usually rely on cohort design → time-to-event data (censoring)

b) **Both current methods use the PS:**
   - PS captures confounder-exposure associations, but **does *not* account for Confounder-Outcome associations**

Our Approach: Imputation of Missing Confounders

General Idea:

We propose to Impute Unmeasured Confounders, for all subjects in the main database, based on information in Validation Subsample regarding their Relationships with both *(i) Exposure* and *(ii) Outcome*

[Burne & Abrahamowicz, *Stat Med* 2016]

Challenge:

How to represent the 'Outcome', which in the context of Survival Analyses, is Bi-Variate, i.e. includes 2 terms:

(1) a Continuous measure of Follow-up Duration +

(2) a Binary indicator of "Status" (Event or Censored) at the end of follow-up?

Our Approach: use **Martingale Residuals to Impute missing confounders** based on VS data

- We propose to **Use Martingale Residuals (MR) in imputation model** [Burne & Abrahamowicz, *Stat Med* 2016]**:**
  - MR's contain info on ***lack of fit ('bad prediction' of individual outcomes based on measured variables)***
  - Such 'bad prediction' may be informative about values of un-measured confounders (U), which, by definition, must be associated with the outcome

** Accounting for the Outcome should improve Imputation [Moons et al, *J Clin Epi* 2006]

- Yet, PS (used in previous methods) only captures relationship with exposure

Rationale for using RESIDUALS

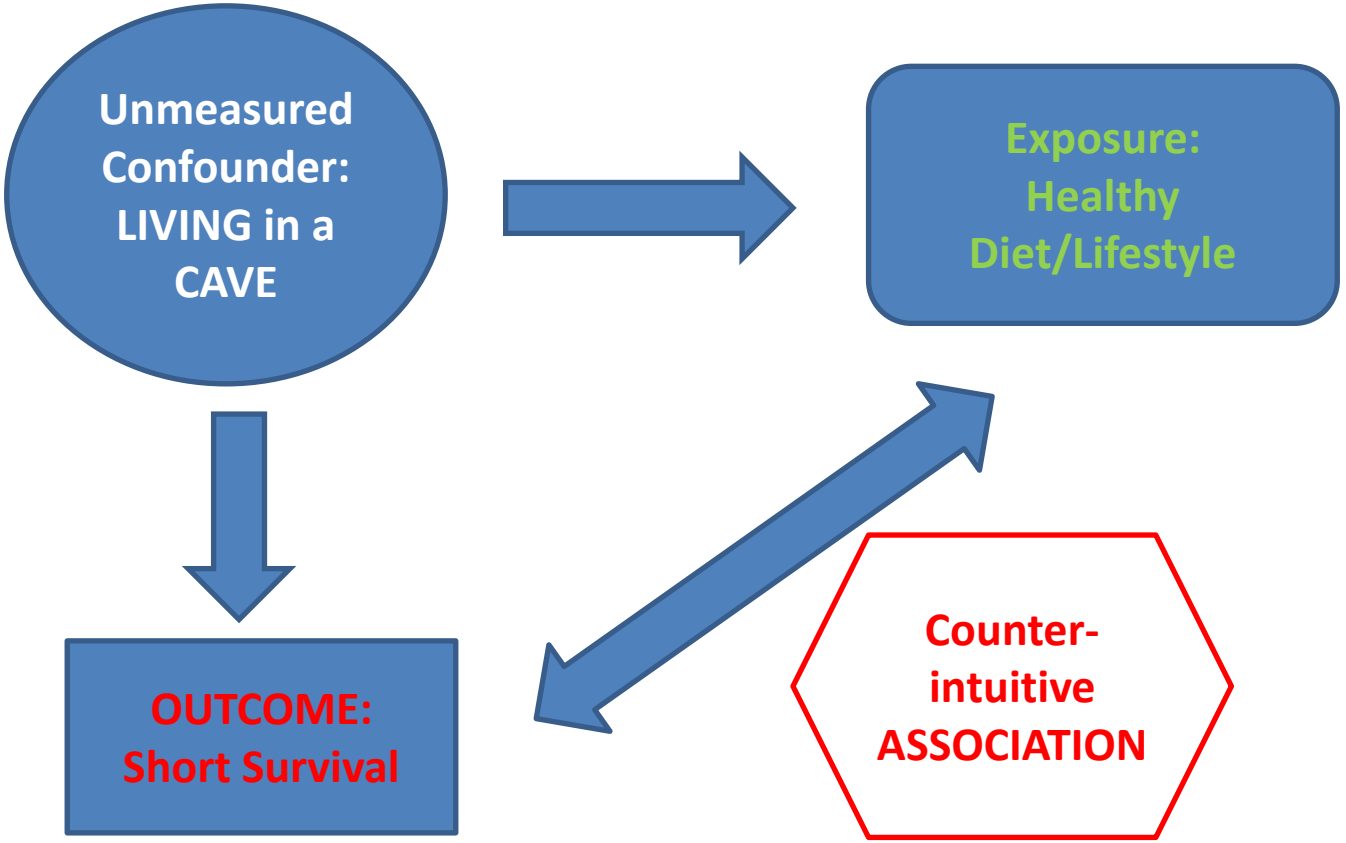
- **Residuals reflect the Discrepancies between:**
  a) **the Outcome Actually Observed for individual patients** ***versus***
  b) **the Outcomes Predicted by the Model** (that adjusts Only for the fully measured confounders)
- Thus, e.g. a Large Positive Residual will suggest that we may be Missing an Important Risk Factor (that could 'explain' why the patient had an unexpectedly Bad Outcome etc.)

An **Early Epidemiological Puzzle:**

"*Something does Not feel right*: *our Air is Clean, Water is Pure, we get Plenty of Exercise, everything we Eat is Organic*, ... *and yet Nobody Lives Past 30* "

DAG for the "Cavemen Epi puzzle":

CAVE Living = MISSING CONFOUNDER that is a common cause of Both: Exposure & Outcome and explains the apparent Discrepancy between Observed vs. 'Expected' Outcome!



"Proof of Concept": Martingale Residuals Capture info on Un-measured Confounder

Simulated Data (N ~ 12,000):

- Outcome depends on Exposure (A), 2 measured confounders (X1, X2) and **UN-measured Confounder U** (range from 0 to 1)**, with Higher values of U □High**
- Highest Martingale Residuals observed for subjects with Very Early Events (Shortest Survival) who have "Protective Values" of Exposure (A) and X1, X2:

  Mean value of U for 100 Highest MR = 0.70
- Lowest Martingale Residuals observed for subjects Censored at the Latest Time (Longest Survival) with High-Risk values of A, X1 & X2:

  Mean value of U for 100 Lowest MR = 0.35

Data Structure

- Individual *i*'s data: $\{t_i, \delta_i, X_i, \boldsymbol{C}_i, U_{1i}, \ldots, U_{ki}\}$
- t_i, δ_i : time, indicator for event
- X_i: exposure
- $\boldsymbol{C}_i$: all confounders measured in main database
- For the $i = n+1, \ldots, N$ individuals in *main database*, $U_{1i}, \ldots U_{ki}$ are unmeasured
- For the $i = 1, \ldots, n$ individuals in *validation subsample*, $U_{1i}, \ldots U_{ki}$ are measured

Steps 1-3 of "Martingale Imputation"

In **the main database** (*i* = 1,..., *N*), perform **steps 1-2**:

1. Fit Cox PH model dependent on *X* and ***C***:

$$\lambda_i(t) = \lambda_0(t)\exp\{\gamma_1 X_i + \gamma_2 C_i\}$$

2. Obtain Martingale Residuals M(i) from model in step 1

In **Validation Sample** (*i* = 1,..., *n*) perform **step 3:**

3. Estimate (separately for each *U* in U_1, ..., U_k)
   expected distribution for ***U***, based on $\hat{M}$, *X* and ***C***
   (where $\hat{M}$ is obtained from step 2)

Steps 4-6 of "Martingale Imputation"

In the **main database** (i =1,..., N), perform **steps 4-6**:

4. Impute individual values of each U by sampling from the respective distribution obtained in step 3
5. Re-run Cox model to estimate effect of X adjusted for $\boldsymbol{C}$ and imputed values of U_1,..., U_k
6. Use bootstrap to obtain 95% CI's for HR for exposure (X) from the model fit in step 5

R program for MR Imputation

- Current Link for the code used for the Simulations in the [Burne & Abrahamowicz, *Stat Med* 2016] paper: http://github.com/RMBurne/MR-based-imputation
- Users familiar with R programming can adapt parts of the code to apply the method for analyzing real-life datasets
- Alternatively, contact the authors for a new link (to be created in the next 2-3 weeks) for the more user-friendly code to implement the method (and example of an implementation)

Simulation Objectives

1. **Validate the proposed MR-based imputation**, for different simulated data structures, in terms of:
   i. Lack of BIAS of the estimated log HR for 'exposure' &
   ii. Correct Coverage Rate of the bootstrap-based 95% CI's
2. Compare the Bias, SD & RMSE of the MR-based estimates with the results of:
   i. Conventional Cox model (adjusted only for ***C***)
   ii. "Standard" imputation of ***U*** (based on ***X*** and ***C*** but not the outcome)
   iii. PSC (while Varying the Surrogacy assumption)

Simulation Design: basic set-up

- Full data (main database + validation subsample) N = 10,000
- Validation subsample (VS) n = 1,000 (10%)
- Cumulative incidence of event = 10% (90% censoring)
- Binary (time-fixed) exposure
- 2 confounders (C_1, C_2) measured in the main database
- 2 additional confounders (U_1, U_2) measured only in validation subsample

Simulation Design: ‘sensitivity analyses’

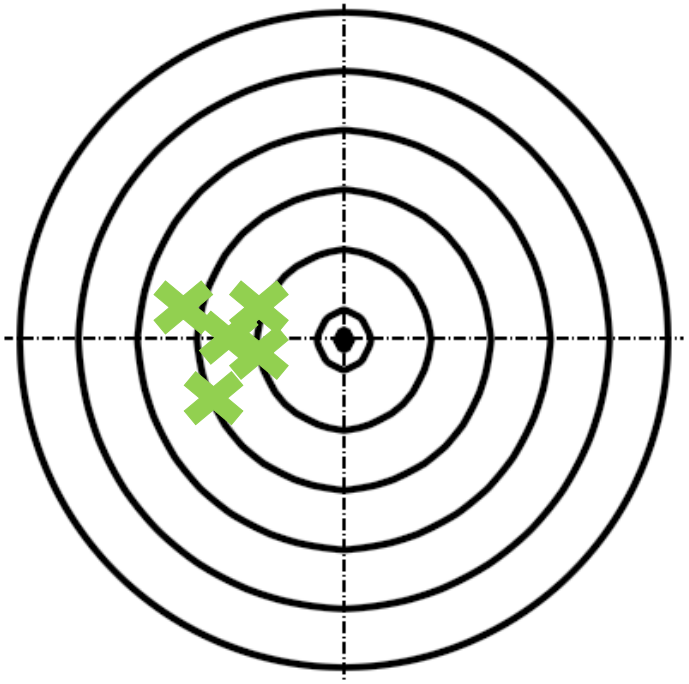
- 1,000 replicates per simulated scenario
- Across different simulated scenarios we varied:
  i. Strength & direction of unmeasured confounding
  ii. Relationship of *U* and *C* with outcome and exposure so surrogacy either held or was violated to different extents
  iii. True HR for exposure
  iv. Censoring mechanism
  v. Validation subsample size (n=1,000, 500 or 250)

Comparison of BIAS in the exposure HR estimated with 4 models

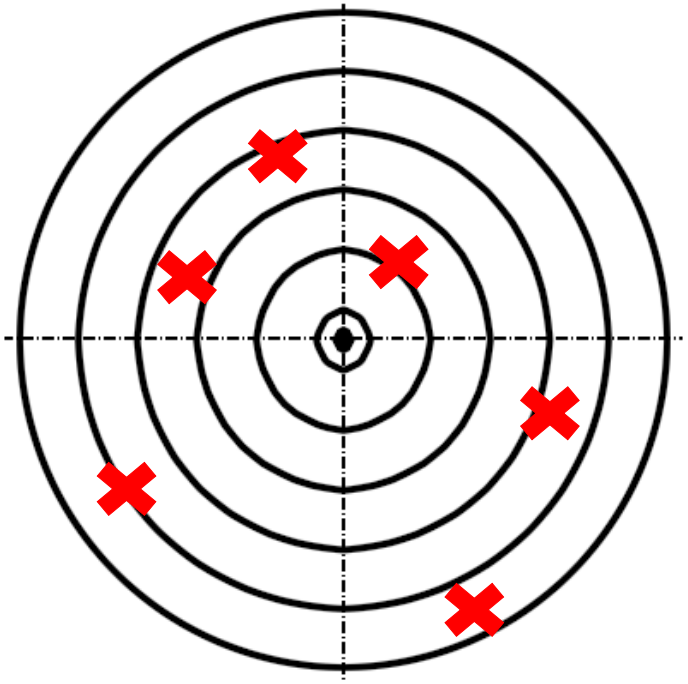
[n(VS) =1,000, moderate unmeasured confounding, surrogacy satisfied]

	True HR	Martingale Imputation	Imputation without Martingale	Conventional Cox PH	PSC
1	1	**0.001**	0.144	0.160	-0.012
2	1.2	**0.004**	0.147	0.164	-0.009
3	1.5	**0.003**	0.143	0.159	-0.013

Mean Squared Error (MSE) = $(Bias)^2$ + Variance
compares the **Overall Accuracy of the Estimates**



Moderate BIAS + Low Variance = Low RMSE

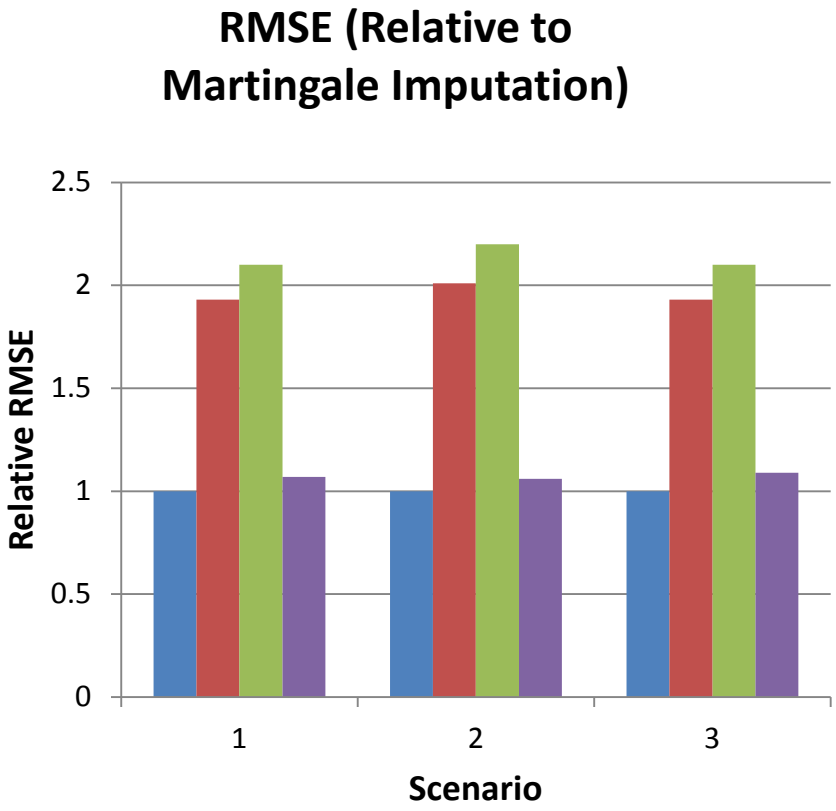


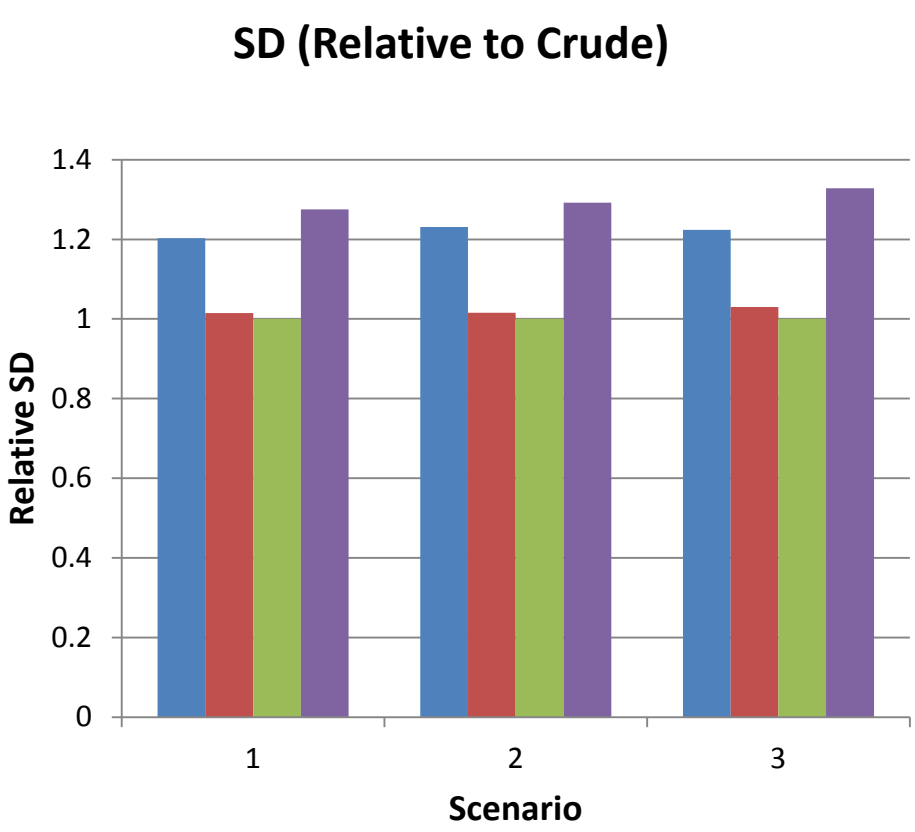
NO Bias but High VARIANCE = High MSE

(Root MSE) MSE = √ MSE (Lower RMSE = Better Accuracy)

RMSE & SD of exposure HR estimated with 4 methods

[n(VS) =1,000, moderate unmeasured confounding, **surrogacy satisfied**]





Martingale Imputation Imput. Without Martingale Conventional PSC

Comparison of BIAS for various strengths of unmeasured confounding & random censoring

[n(VS) =1,000, true HR = 1, **surrogacy satisfied**]

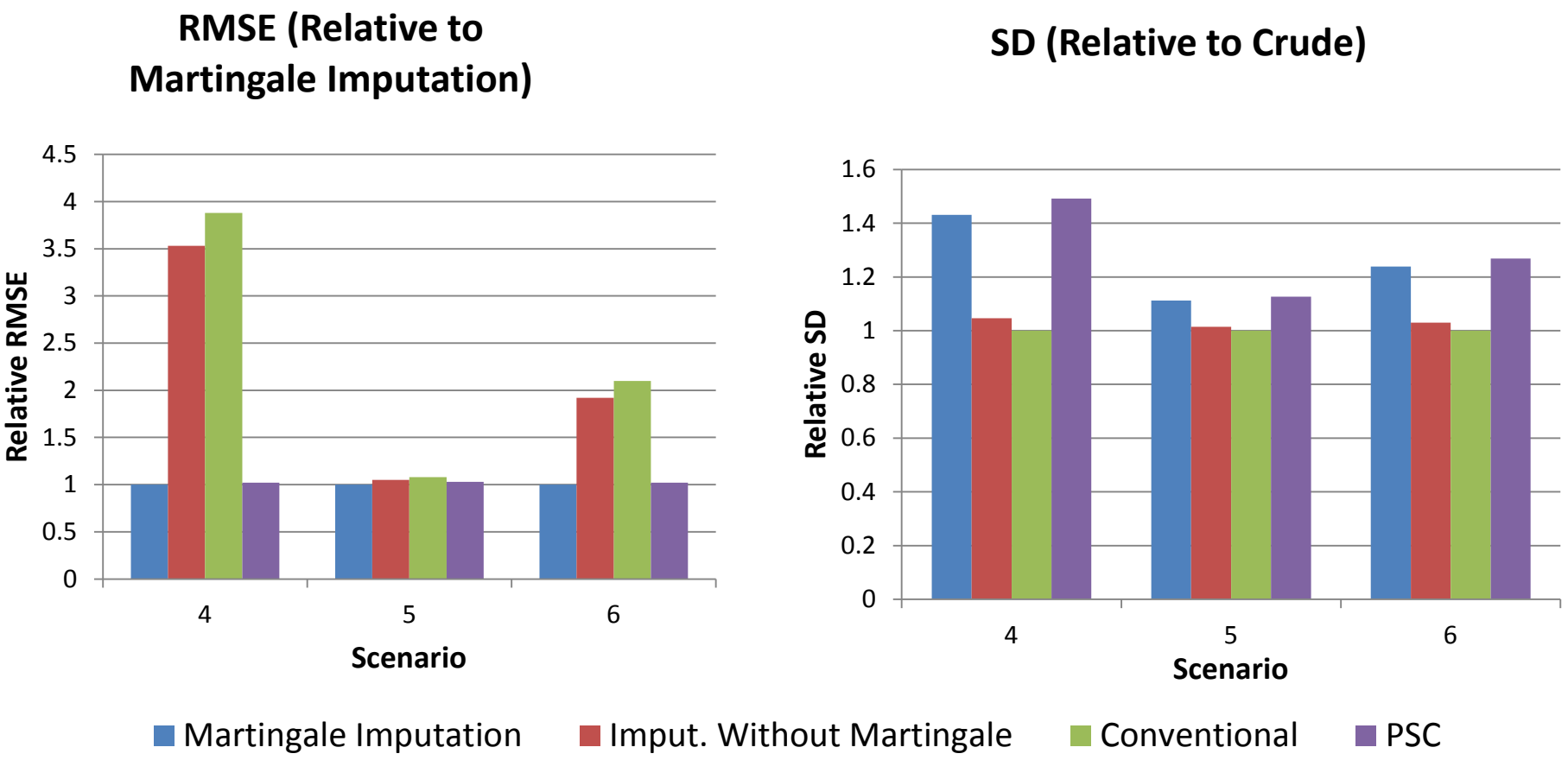
	Alter	Martingale Imputation	Imputation without Martingale	Conventional Cox PH	PSC
4	Strong unmeas. conf.	0.020	0.328	0.363	-0.007
5	Weak unmeas. conf.	-0.001	0.041	0.046	-0.013
6	Censoring mech.	0.003	0.146	0.162	-0.009

Coverage of the 95% Bootstrap CI's

(selected scenarios, 100 events in VS, Surrogacy satisfied)

	Martingale Imputation	Imputation without Martingale	Conventional Cox PH	PSC
1	94.8	46.1	37.4	93.5
2	95.2	38.2	27.9	94.5
3	94.8	41.3	32.2	93.8
4	93.2	0.1	0.0	94.5
5	94.4	90.0	89.3	94.6
6	95.6	42.9	32.7	95.9

Comparison of RMSE & SD for various strengths of unmeasured confounding (surrogacy satisfied)



4 = strong conf., 5 = weak conf., 6 = random censoring

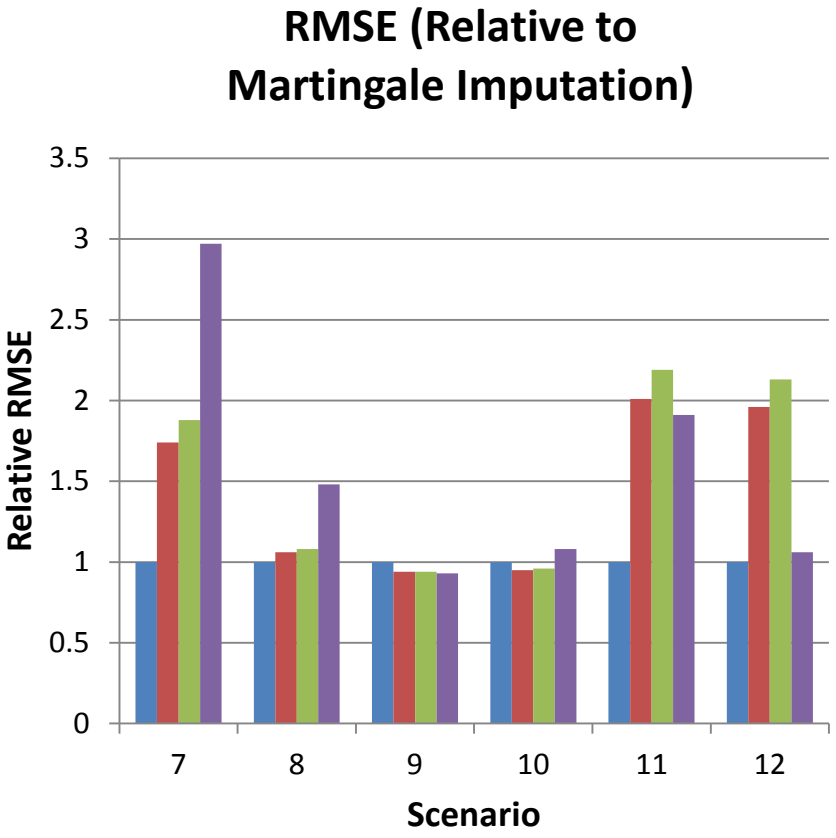
Surrogacy Violated: comparison of BIAS of log(HR) for Martingale Imputation-based vs. PSC estimates

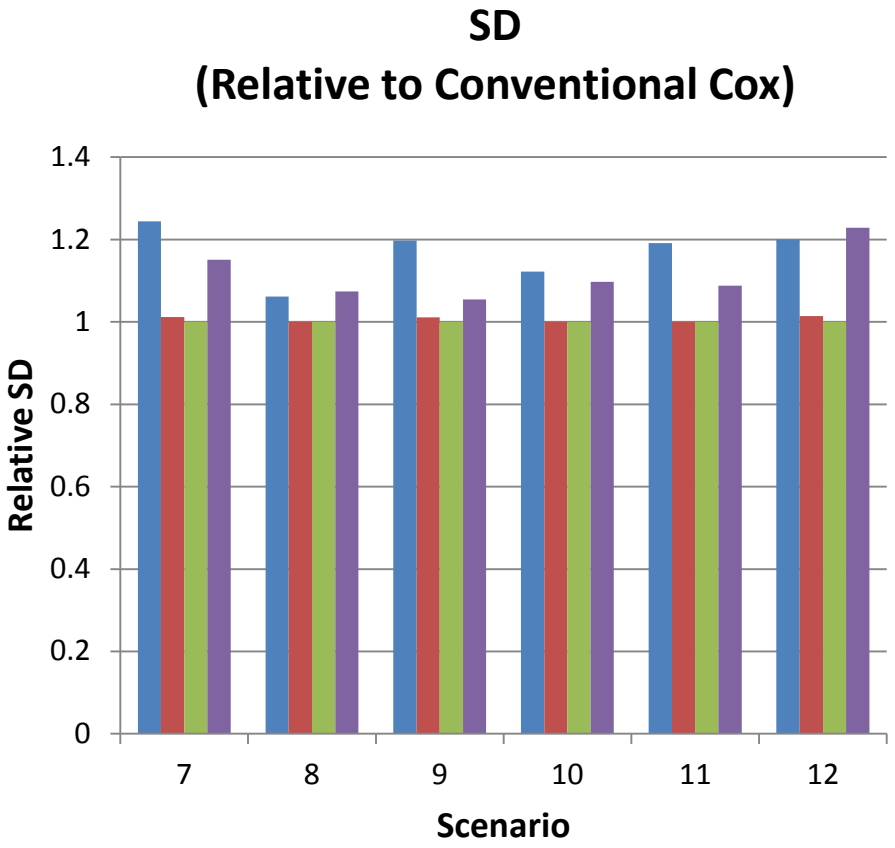
[n(VS) =1,000, ~100 events, true log HR = 0, weak to moderate confounding by U]

	Scenario		Martingale Imputation	PSC	Results of SURROGACY Test	
	Surr. viol.	Unmeas. conf.	Bias	Bias	Mean Surr. LR	% Surr. pval < 0.05
7	Strong	Moderate	-0.003	-0.303	<0.001	100.0
8	Moderate	Weak	-0.003	-0.092	3.50	94.0
9	Weak	Moderate	-0.003	-0.030	54.57	20.1
10	Moderate	Moder/Weak	-0.001	-0.041	35.57	43.3
11	Strong	Moderate	0.005	0.137	3.50	93.2
12	Weak	Moderate	0.007	0.024	55.38	18.2

Impact of Surrogacy Violation ** on RMSE & SD

(** sc. 7 & 11: Strong; sc. 8 & 10: Moderate, sc. 9 & 12: Weak)





Martingale Imputation Imput. Without Martingale Conventional PSC

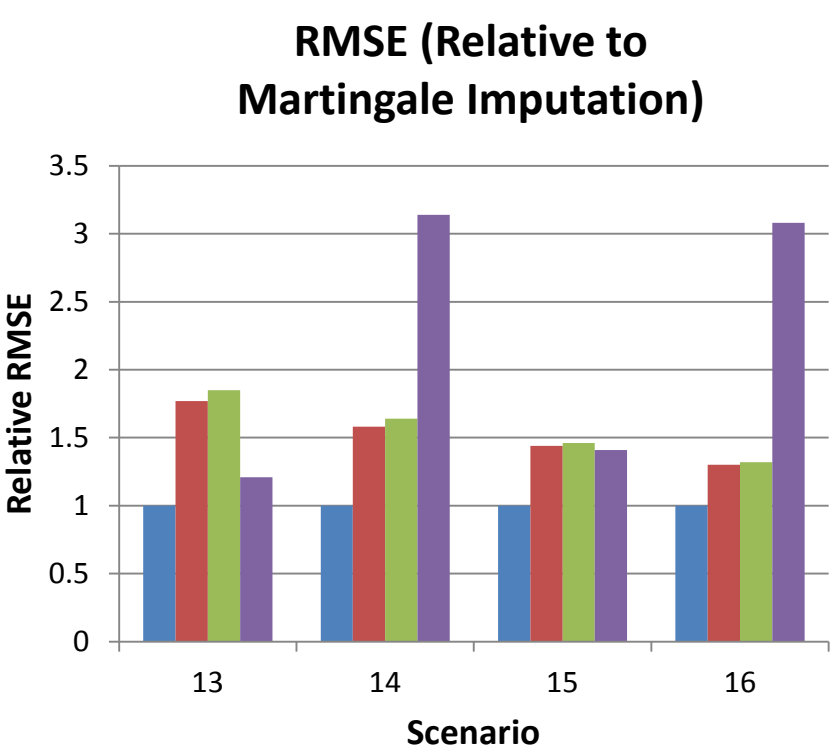
BIAS comparisons for different **validation subsample size**

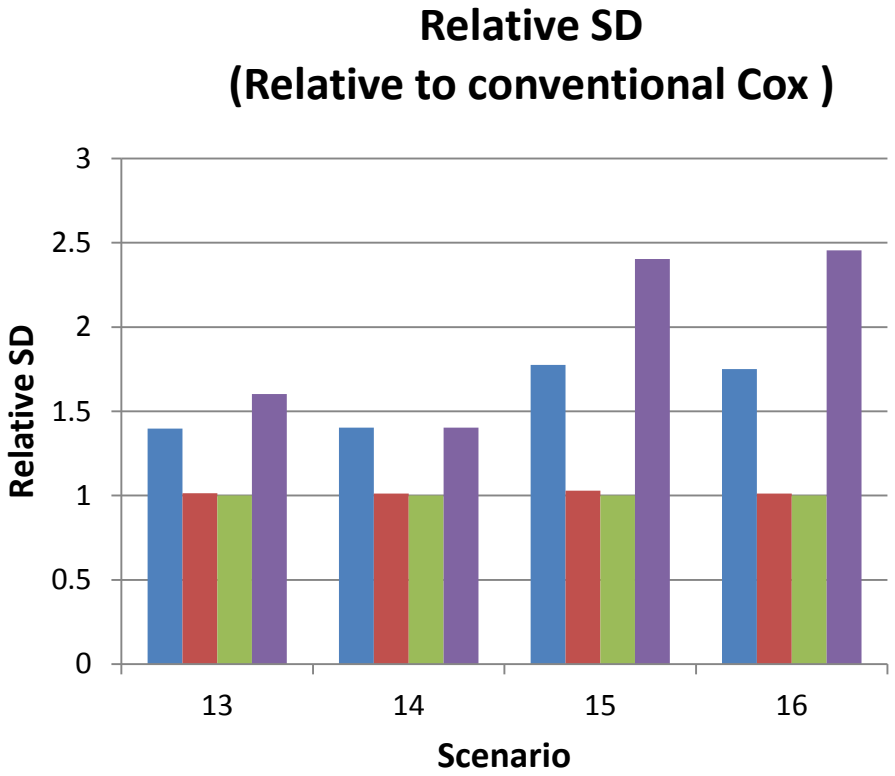
[True HR = 1, Moderate unmeasured confounding]

	n (VS)	Surrogacy	Martingale Imputation	Imputation without Martingale	Conventional Cox PH	PSC
13	**500**	**Met**	**0.000**	0.154	0.162	**-0.036**
14	**500**	**Violated**	**0.002**	-0.171	-0.180	**-0.363**
15	**250**	**Met**	**0.010**	0.156	0.161	**-0.047**
16	**250**	**Violated**	**-0.004**	-0.180	-0.184	**-0.425**

RMSE & SD Comparisons for different **N** of the **Validation Subsample (VS)**

(<u>50 events in VS</u> in sc. 13 & 15, OR <u>25 events</u> in sc. 14 & 16)



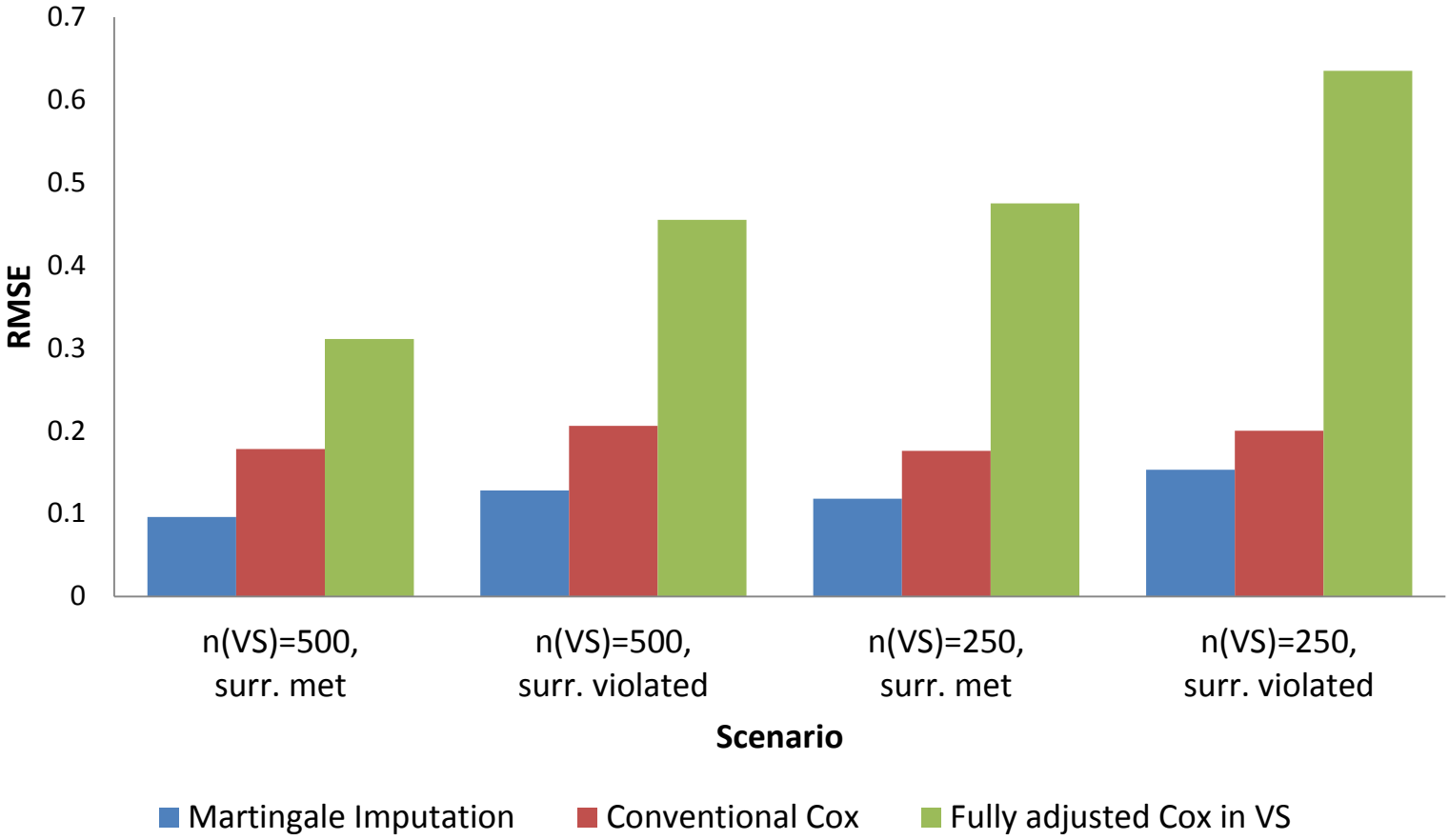


Martingale Imputation Imput. Without Martingale Conventional PSC

RMSE of exposure estimates:

MR-imputation **vs**

Using ONLY VS data with (directly adjusted for U)



Summary of Simulation Results

- **Proposed Martingale-based Imputation**:
  i. **UN-BIASED estimates** for All Scenarios considered
  ii. **Usually the Best Overall Accuracy (lowest RMSE** in Almost All Scenarios)
  iii. **Performed as well as Propensity Scores Calibration (PSC) IF the Surrogacy assumption was met** or only slightly violated
  iv. **Performed well even IF Surrogacy was Violated** (in contrast to Large Bias & high RMSE's of PSC estimates)
  v. **Consistently Improved Accuracy over standard Imputation** based on U-X associations only
  vi. **Performed well even if Validation Subsample included as few as 25-50 Events**

Application # 1:
use Glucocorticoids vs. type II Diabetes

BACKGROUND:

- Glucocorticoid (GC) therapy: common treatment for rheumatoid arthritis (RA)
- Very effective in slowing disease progression
- Potentially serious side effects, including diabetes mellitus type II
- Published results of the GC – DM are ambiguous

[Burne & Abrahamowicz, *Stat Med* 2016]

Data: Main Database

- **CPRD (Clinical Practice Research Datalink)**:
  - **RA patients in the UK**, identified by a validated algorithm
  - Includes patient demographics, medical diagnoses, drug prescriptions...
  - **Some potential confounders, including BMI, disability level, comorbidity index, not available**
  - N = 16,898
  - # events = 1,665 (10%)
  - Mean follow-up = 6.5 years

[Movahedi et al, *Arthritis & Rheumatology* 2016]

External Validation Sample (VS)

- **VS = NDB**
  (National Data Bank for Rheumatic Diseases):
  - Longitudinal observational study of **patients with RA in the US**, created for research purposes
  - **More rich confounder information, including BMI, disease severity (HAQ score), comorbidity**
  - Also includes all those variables available in CPRD
  - N = 8,253
  - # events = 529 (6%)
  - Mean follow-up = 4.5 years

[Wolfe and Michaud, *Rheumatology* 2011]

Analyses

- Time-to-Event analyses based on Multivariable Cox proportional hazards (PH) model
- Time 0 = RA diagnosis
- Event time = diagnosis of type II DM
- Time-varying ** Exposure = binary indicator of the Current Use of GC
- 2 Models in the CPRD database:
  - Conventional Cox's model (adjusted only for confounders measured in the CPRD)
  - Cox model with MR-imputation of additional confounders (measured only in the NDB database)

Comparison of the NDB vs. CPRD populations of RA patients

Characteristics of the main database (CPRD) and validation data (NDB)

	NDB ($n = 8253$)	CPRD ($n = 16898$)
Incident type II diabetes mellitus	529 (6.4%)	1665 (9.9%)
Follow-up (days): mean (SD)	1670.0 (1359.0)	2364.8 (1643.5)
Event rate (/100 person-years) (95% CI)*	1.40 (1.28, 1.53)	1.52 (1.45, 1.60)
Used glucocorticoid during follow-up: *n* (%)	2358 (28.6%)	5864 (34.7%)
Days exposed amongst users: median (IQR)	304.0 (92.0–732.8)	219.0 (36.0–831.2)
Incidence rate of new exposure (/100 person-years) (95% CI)*	8.34 (8.00, 8.68)	7.22 (7.04, 7.41)
Mean % time exposed since first exposure	50.75%	38.75%
Sex = male: *n* (%)	1554.0 (18.8%)	4953.0 (29.3%)
Baseline age: mean (SD)	58.6 (13.3)	58.1 (14.6)
History of NSAID use at cohort entry: *n* (%)	6126 (74.2 %)	14335 (84.8%)
Methotrexate use in follow-up: *n* (%)	4280 (51.9%)	3530 (20.9%)
Hydroxychloroquine use in follow-up: *n* (%)	2329 (28.2%)	1110 (6.6%)
Baseline HAQ disability score: mean (SD)	0.905 (0.697)	–
Baseline comorbidity index: median (IQR)	1 (0–2)	–
Baseline BMI: mean (SD)	27.89 (6.42)	–

Results: NDB Models

	Reduced model		Full model	
	HR	**95% CI**	**HR**	**95% CI**
Glucocorticoid use*	1.81	(1.45, 2.27)	1.48	(1.17, 1.86)
Sex (Male = 1)	0.91	(0.73, 1.15)	1.00	(0.79, 1.27)
Baseline Age	0.99	(0.99, 1.00)	0.99	(0.99, 1.00)
NSAID use (before cohort)	0.73	(0.59, 0.91)	0.73	(0.59, 0.90)
Methotrexate use (TD)	0.81	(0.68, 0.96)	0.89	(0.75, 1.06)
Hydroxychoroquine use (TD)	0.70	(0.57, 0.87)	0.73	(0.59, 0.90)
HAQ disability score (TD)			1.15	(1.01, 1.31)
Comorbidity index (TD)			1.24	(1.17, 1.31)
BMI (TD)			1.18	(1.10, 1.26)
BMI2 (per 50 units) (TD)			0.93	(0.89, 0.98)

* Current GC use modelled as a time-dependent variable.

PRIMARY Results: CPRD analyses

	Reduced model		Full model	
	HR	**95% CI**	**HR**	**95% CI**
Glucocorticoid use*	**1.38**	(1.19, 1.59)	**1.15**	(0.99, 1.33)
Sex (Male = 1)	1.29	(1.16, 1.43)	1.56	(1.40, 1.73)
Baseline Age	1.02	(1.01, 1.02)	1.02	(1.01, 1.02)
NSAID use (before cohort)	1.07	(0.94, 1.22)	1.15	(1.01, 1.31)
Methotrexate use (TD)	1.22	(1.06, 1.40)	1.28	(1.12, 1.48)
Hydroxychoroquine use (TD)	0.86	(0.64, 1.15)	0.91	(0.68, 1.22)
HAQ disability score (TD)			1.52	(1.41, 1.64)
Comorbidity index (TD)			1.36	(1.31, 1.41)
BMI (TD)			1.28	(1.21, 1.35)
BMI2 (per 50 units) (TD)			0.89	(0.86, 0.92)

* Current GC use modelled as a time-dependent variable.

Summary of the Results

- NDB analyses indicated that several variables NOT measured in the main (CPRD) database are significant risk factors for DM II and that adjusting for these variables may attenuate considerably the estimated impact of current GC exposure
- Consistent with these results, application of the proposed MR-imputation in the CPRD analyses resulted in an important reduction of the estimated effect of GC: from **1.38** (1.19,1.59) to a (marginally non-significant) **1.15** (0.99, 1.33)

Next Step: extension to Marginal Structural Models

- **Marginal Structural Models (MSMs)** are increasingly used to **account for Time-Varying covariates that may both Confound and Mediate the effects of Treatment or Exposure**
  [Hernan et al, *Epidemiology* 2000]

- MSM use **Inverse Probability of Treatment Weights (IPTW)** to re-weight observations so as to create a `randomized' pseudo-population, in which covariates are balanced across the treatment groups

- MSM require **Assumption of NO Unmeasured Confounders**

- No validated methods exist to deal with unmeasured confounding in MSM**

** Exceptions, developed for Specific Data Structures:
[Brumback et al, *Stat Med* 2004; Moodie et al, *Int J Biostat* 2008]

Proposed MR-based MSM methods

2 alternative methods, both use Martingale Residuals (MR)
(see previous slides)
to account for Time-Varying Confounder U(t) Measured Only in the VS, **when estimating IPTW** [Burne and Abrahamowicz, *Stat Methods Med Res* 2017]

1) MR-based Multiple Imputation:

   step 1: use MR's to impute individual values of U(t)

   step 2: use MR-imputed values to estimate IPTW

2) MR-enhanced Regression Calibration of the PS:

   Method akin to propensity score calibration (PSC)* but uses Martingale Residuals in the regression calibration model

* [Sturmer et al, *AJE* 2005]

Comparator methods

3) Regression calibration of the PS:
   - Method akin to propensity score calibration (PSC)*
   - Use regression calibration to correct PS in main data, using data in VS
   - Corrected PS used in calculation of IPTW weights

4) 'Naive model':

   ignores *U(t)* (NOT measured in the main database) while estimating IPTW

* [Sturmer et al, *AJE* 2005]

Simulation objectives

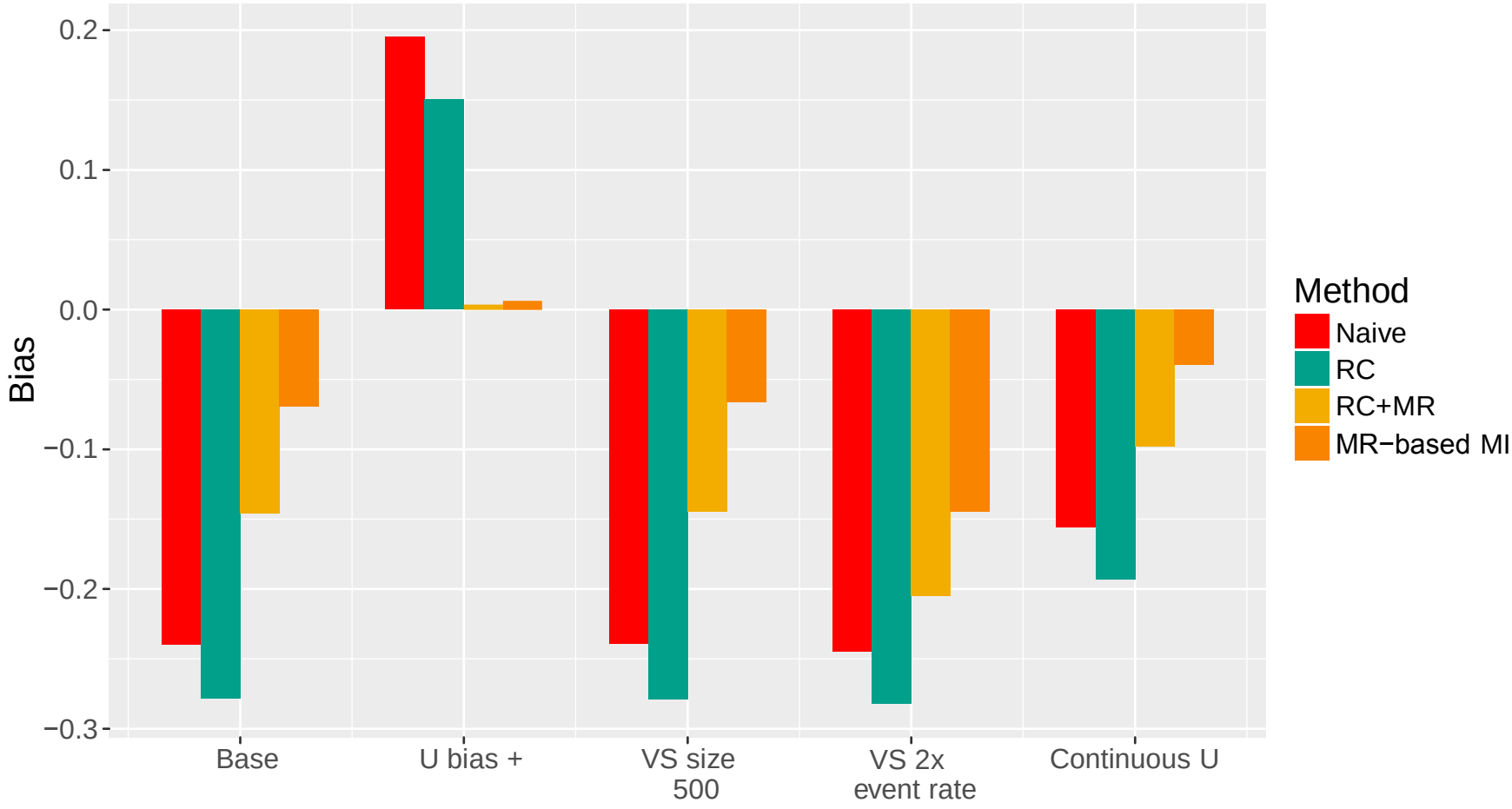
- To compare performance of 4 alternative estimators:
  - Naive MSM (IPTW not account for Unmeasured Confounder)
  - Regression calibration (RC)
  - MR-enhanced regression calibration (RC + MR)
  - Martingale residual-based multiple imputation (MR-based MI)
- With respect to: Bias, standard deviation (SD), root mean squared error (RMSE)

[Burne and Abrahamowicz, *Stat Methods Med Res* 2017]

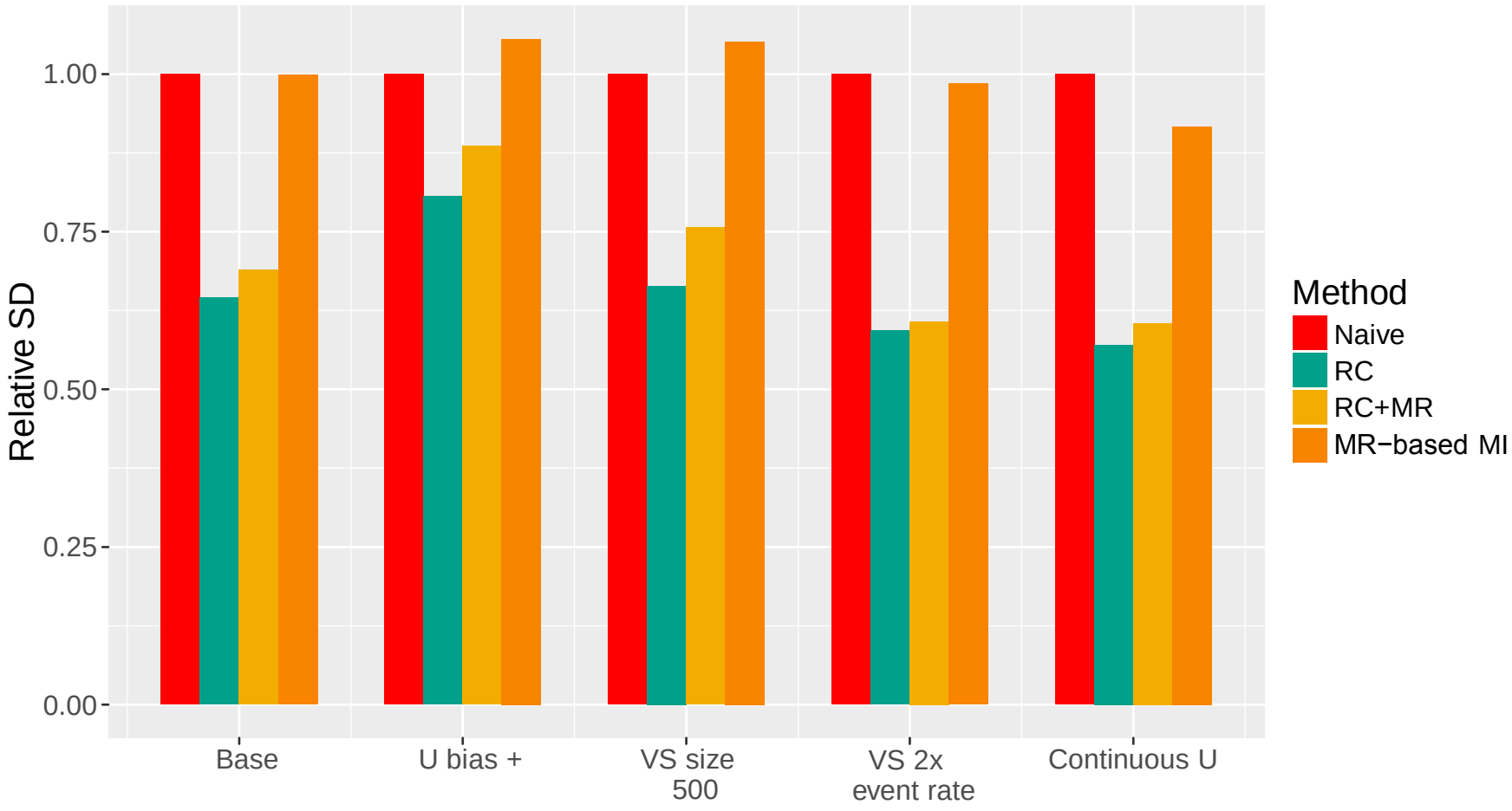
Simulation Design

- Hypothetical cohort study with up to 10 visits
- Binary, time-varying exposure A(t)
- 4 measured confounders (2 baseline, 2 time-varying)
- 1 unmeasured time-varying confounder U(t)
  (available in VS Only)
- Outcome data generated based on method by Young et al. [Lifetime Data Analysis 2010]
- N_{main} = 10,000, n_{VS} = 1,000
- 1,000 replications

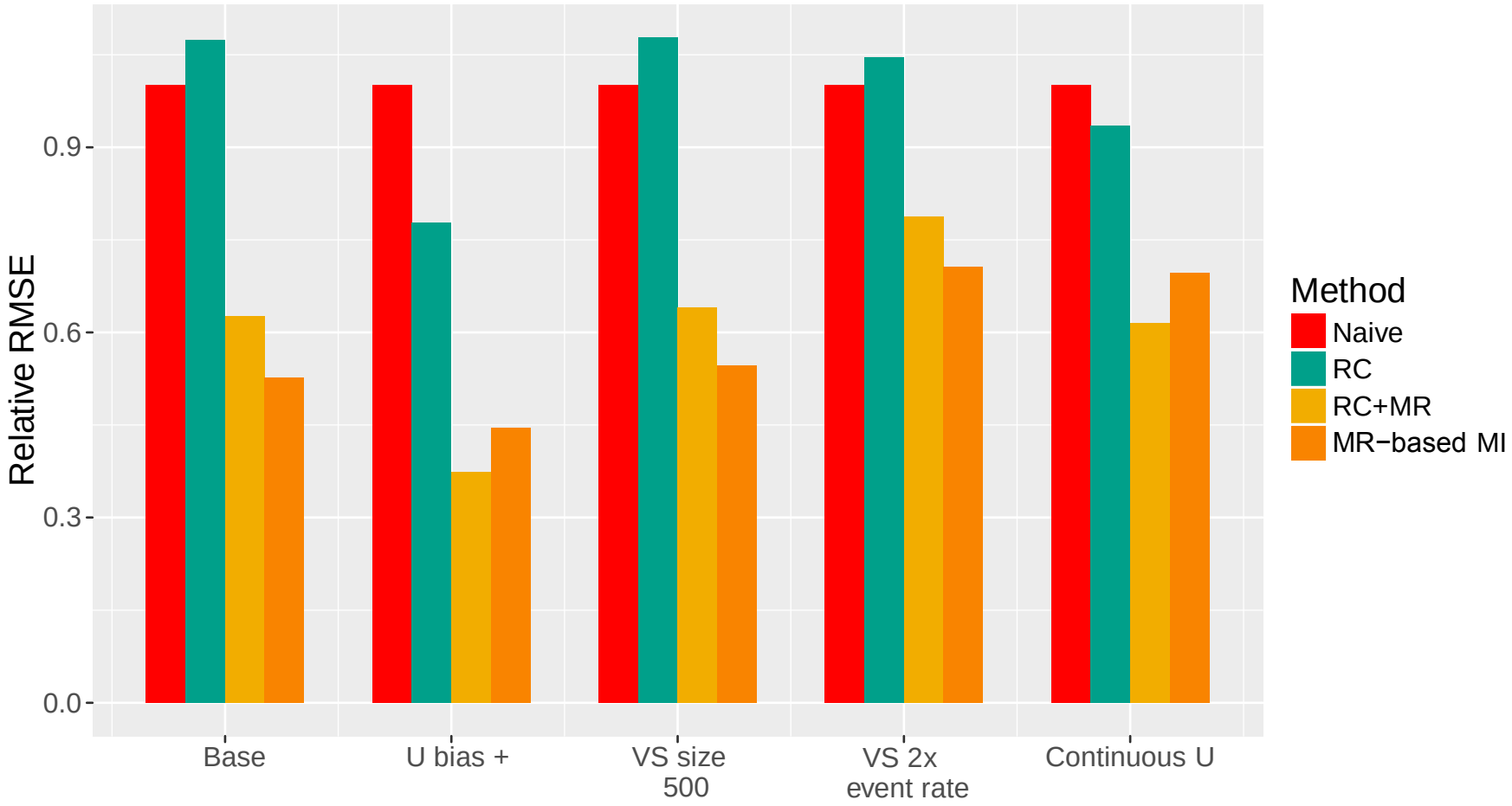
Simulation results: Bias



Simulation results: SD



Simulation results: RMSE



Summary of MSM Simulation Results

- **MR-based MI eliminated Bias and yielded more accurate estimates (lowest RMSE) than the other methods**
- RC of PS sometimes increased bias
- Including martingale residual enhanced RC method
- MR-based MI had larger variation than RC methods, but lowest RMSE (greatest overall accuracy)

2nd Application: MSM analyses

- **Does persistence with DPP-4 inhibitor therapy reduce risk of hospitalization due to hypoglycemia in DM II?**
- Cohort of diabetic patients from Truven Health MarketScan
- Inclusion criteria (for this application):
  - ≥ 1 ICD-9 code for type II Diabetes
  - initiating DPP-4 inhibitors (no prescription in prior year)
  - as third-line therapy (prior prescription to metformin & sulfonylurea)
- 50,305 patients met criteria (Main Database)
- Of those, 2,341 had ≥ 1 measurement of HbA1c available ('internal' Validation Sample)

Analyses

- MSM Cox PH models with IPTW
- Time 0 = 1st prescription of DPP-4 inhibitors between 1 January 2011 and 30 December 2014
- Event = 1st Hospitalization for Hypoglycemia
- Time-Varying Exposure = Current Use of **DPP-4 inhibitors**
- **Time-varying Confounder/Mediator (used to estimate IPTW), measured only in a small subset (VS) = updated value of HbA1c** (important determinant of the initiation of the DPP-4 treatment)
- Baseline covariates: age, sex, employment status, Charlson Comorbidity Index, and healthcare services utilization, in a 1-year period before the cohort entry

Cohort characteristics

	Main Database (N = 47,964)	Validation Sample (n = 2,341 with HbA1c)
Event (hypoglycemia)	2867 (6%)	130 (5.6%)
Follow-up (days): mean (SD)	591.9 (360.9)	592.9 (372.4)
Days exp. to DPP-4: mean (SD)	248.9 (231.3)	216.3 (199.7)
Baseline characteristics:		
Age (years): mean (SD)	58 (11.1)	56.5 (10.1)
Charlson index: mean (SD)	0.4 (0.9)	0.4 (0.9)
Female: n (%)	19348 (40.3%)	1011 (43.2%)
Employed: n (%)	17875 (37.3%)	570 (24.3%)
Characteristics in year prior to cohort entry:		
Emergency department visits: n (%)	10000 (20.8%)	539 (23%)
Hospitalizations: n (%)	4528 (9.4%)	175 (7.5%)
$\geq$ 20 physician visits: n (%)	5495 (11.5%)	223 (9.5%)
Prior hypoglycemic event: n (%)	1925 (4%)	113 (4.8%)
Time-varying characteristics:		
Days exp. to other anti-diabetic: mean (SD)	475.6 (361)	458.8 (365.2)
Baseline HbA1c: mean (SD)	-	8.7 (1.7)
No. HbA1c tests: mean (SD)	-	2.3 (1.8)

Results: Adjusted HR for current exposure to DPP-4 inhibitors

Method	HR	95% CI
Naive	0.62	(0.47, 0.82)
Regression Calibration	0.69	(0.59, 0.82)
Regression Calibration + MR	0.69	(0.59, 0.82)
MR-based multiple imputation	0.74	(0.57, 0.97)
Validation sample*	0.84	(0.55, 1.27)

Weights truncated at 99.7th percentile

* Weight accounts for most recent HbA1c value

Summary of Results

- Minor impact of correcting IPTW for HbA1c:

  HbA1c is a strong determinant of the DPP-4 Treatment but NOT associated with DM II risk

- MR-based MI results confirmed conventional estimates, but reduced concern of unmeasured confounding by HbA1c

- “Complete case” analysis (limited to VS) not adequate, low power, Wide Confidence Intervals that include HR=1 (□ uncertainty: is there an association at all ?)

Overall Conclusions

- Efficient use of additional measurements available in smaller Clinical Validation Subsamples may avoid Bias due to Un-measured Confounding and substantially improve the overall accuracy of the estimates
- Proposed Martingale Residual (MR)-based Imputation provides a promising method for dealing with this important challenge in Survival Analysis and outperforms existing methods such as PSC
- Applications confirm that MR-based imputation may substantially changed the estimated strength of the Drug use association with Adverse Events (e.g. GC vs. DM II)

(selected) **Directions for Future Research**

- Further **simulations** with More Un-measured Confounders (& **different assumptions about**
  U-X, U-Y, U-C associations)
- **Real-life Applications**
- Need to assess the performance of the MR-based imputation if **Validation Subsample has different characteristics that the Main Database ?**

- THANK YOU

- Michal.Abrahamowicz@McGill.ca

McGill

CAN-AIM

References

- Abrahamowicz M, Beauchamp M-E, Ionescu-Ittu R, et al. Reducing the Variance of the Prescribing Preference-Based Instrumental Variable Estimates of the Treatment Effect. *Am J Epidemiology* 2011; 174(4): 494-502.
- Abrahamowicz M, Bjerre L, Beauchamp M-E, LeLorier J, Burne R. The Missing Cause Approach to Unmeasured Confounding in Pharmacoepidemiology. *Stat Med* 2016; 35(7): 1001-1016.
- Avorn J. In defense of pharmacoepidemiology - embracing the yin and yang of drug research. *N Engl J Med* 2007, 357: 2219.
- Brookhart MA, Wang PS, Solomon DH, et al. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006; 17(3): 268-275.
- Brumback BA, Hernan MA, Haneuse SJ, et al. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med* 2004; 23: 749–767.
- **Burne R, Abrahamowicz M. Martingale residual-based method to control for confounders measured only in a validation sample in time-to-event analysis*. Stat Med* 2016; 35(25): 4588-4606.**
- **Burne R, Abrahamowicz M. Adjustment for time-dependent unmeasured confounders in marginal structural Cox models using validation sample data. *Stat Methods Med Res* 2017*;* doi: 10.1177/0962280217726800 [Epub ahead of print].**
- Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol* 1996; 25: 1107-1116.
- Groenwold RH, Nelson DB, Nichol KL, et al. Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research. *Int J Epidemiol* 2010; 39(1): 107-117.
- Hernan MA, Brumback B and Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; 11: 561–570.
- Ionescu-Ittu R, Delaney JAC, Abrahamowicz M. Bias-variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. Pharmacoepidemiology and Drug Safety. 2009; 18(7): 562-571.

References (CONT_D)

- Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol* 2012; 65(2):155-162.
- Kumamaru H, Gagne JJ, Glynn RJ, et al. Comparison of high-dimensional confounder summary scores in comparative studies of newly marketed medications. *J Clin Epidemiol* 2016; 76: 200-208.
- McCandless, L. C., Richardson, S., & Best, N. (2012). Adjustment for missing confounders using external validation data and propensity scores. *J Am St Assoc; 107*(497): 40-51.
- McMahon AD. Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiol Drug Saf* 2003; 12(7): 551-558.
- Moodie EE, Delaney JA, Lefebvre G, Platt RW. Missing confounding data in marginal structural models: a comparison of inverse probability weighting and multiple imputation. *Int J Biostat* 2008; 4: 1–23.
- Movahedi M, Beauchamp M-E, Abrahamowicz M, Ray DW, Michaud K, Pedro S, Dixon WG. Risk of incident type II diabetes associated with dose and duration of oral glucocorticoid therapy in patients with rheumatoid arthritis. *Arthritis Rheumatol* 2016; 68(5): 1089-1098.
- Moons KG, Donders RA, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* 2006*;* 59(10): 1092-1101.
- Patorno E, Garry EM, Patrick AR, Schneeweiss S, Gillet VG, Zorina O, Bartels DB, Seeger JD. Addressing limitations in observational studies of the association between glucose-lowering medications and all-cause mortality: a review. *Pharmacoepidemiol Drug Saf* 2015, 38(3): 295-310.
- Pilote L, Abrahamowicz M, Rodrigues E, Eisenberg MJ, Rahme E. Mortality rates in elderly patients who take different angiotensin-converting enzyme inhibitors after acute myocardial infarction: a class effect? *Ann Intern Med* 2004; 141(2):102-112.

References (CONT_D)

- Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 2009; 20(4): 512-522.
- Skegg DCG. Evaluating the safety of medicines, with particular reference to contraception. *Stat Med* 2001; 20(23): 3557-3569.
- Stürmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol* 2005, *162*(3), 279-289.
- Stürmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration — a simulation study. *Am J Epidemiol* 2007, *165*(10), 1110-1118.
- Walker AM. Confounding by indication. *Epidemiology* 1996; 7(4): 335-336.
- Wolfe F, Flowers N, Burke TA, Arquelle LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity: among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication. *J Rheumatol* 2002; 29(5): 1015-1022.
- Wolfe F, Michaud K. The National Data Bank for rheumatic diseases: a multi-registry rheumatic disease data bank. *Rheumatology* 2011, 50(1): 16-24.